

VŠB – Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky

Katedra kybernetiky a biomedicínského inženýrství

Rozpoznávací systém pro detekci mluveného signálu
Interactive System with Speech Signal Detection

VŠB - Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra kybernetiky a biomedicínského inženýrství

Zadání diplomové práce

Student: **Bc. Ludmila Holušová**
Studijní program: N2649 Elektrotechnika
Studijní obor: 3901T009 Biomedicínské inženýrství
Téma: **Rozpoznávací systém pro detekci mluveného signálu**
Interactive System with Speech Signal Detection

Zásady pro vypracování:

1. Rozbor problematiky systémů pro detekci mluveného signálu.
2. Návrh systému pro detekci mluveného signálu.
3. Realizace systému pro detekci mluveného signálu.
4. Vizualizace a srovnání naměřených výsledků s teoretickými předpoklady.
5. Zhodnocení dosažených výsledků práce.

Seznam doporučené odborné literatury:

- [1] BROUGHTON, S. Allen a Kurt M. BRYAN. *Discrete Fourier analysis and wavelets: applications to signal and image processing*. Hoboken, N.J.: Wiley, c2009, xv, 337 p. ISBN 978-0-470-29466-6.
- [2] CLARK, A., Ch. FOX a S. LAPPIN. *The Handbook of Computational Linguistics and Natural Language Processing*. Malden, MA: Wiley-Blackwell, 2010, xxii, 775 p. ISBN 978-1-4051-5581-6.
- [3] HUANG, X., A. ACERO a H.-W. HON. *Spoken language processing: a guide to theory, algorithm, and system development*. New Jersey: Prentice-Hall, 2001. 980 s. ISBN 0-13-022616-5.
- [4] PSUTKA, Josef. *Komunikace s počítačem mluvenou řečí*. Praha: Academia Praha, 1995. ISBN 80-200-0203-0.
- [5] GOLD, B., N. MORGAN a D. ELLIS. *Speech and Audio Signal Processing*. Hoboken, NJ.: John Wiley & Sons, 2011. ISBN 978-0-470-19536-9.
- [6] PSUTKA, Josef et al. *Mluvíme s počítačem česky*. Praha: Academia Praha, 2006. ISBN-80-200-1309-1.
- [7] NOUZA, Jan. *Počítačové zpracování řeči: cíle, problémy, metody a aplikace: sborník článků*. Liberec: Technická univerzita v Liberci, 2001. ISBN 978-8070835517.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

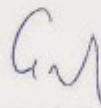
Vedoucí diplomové práce: **Ing. Zdeněk Macháček, Ph.D.**

Datum zadání: 01.09.2014

Datum odevzdání: 07.05.2015



doc. Ing. Jiří Kozíorek, Ph.D.
vedoucí katedry



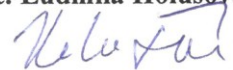
prof. RNDr. Václav Snášel, CSc.
děkan fakulty

Prohlášení

„Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně. Uvedla jsem všechny literární parametry a publikace, ze kterých jsem čerpala.“

Datum odevzdání: 7. 5 2015

Bc. Ludmila Holušová



Poděkování

Mé největší díky patří vedoucímu diplomové práce panu **Ing. Zdeňku Macháčkovi, Ph.D.**, za účinnou metodickou, pedagogickou, odbornou pomoc a další cenné rady při zpracování mé diplomové práce.

Abstrakt

Diplomová práce se zabývá **problematikou zpracování řečového signálu** a rozpoznávání jednotlivých varovných elektronických verbálních sirén – zkouška sirén, požární poplach, nebezpečí zátopové vlny, všeobecná výstraha, chemická havárie, radiační havárie, konec poplachu. Popsána je posloupnost procesů, sloužící k identifikaci dané sirény. V prostředí Matlab je vytvořená softwarová aplikace, která je testována na algoritmem upravených průbězích jednotlivých varovných sirén. Signály jsou modifikovány šumem, posunem, zesílením a zeslabením.

Klíčová slova

Mluvený signál, střední hodnota, energie, elektronické verbální sirény, rozpoznávání verbálních sirén.

Abstract

This thesis deals with **the speech signal processing** and recognition verbal warnings electronic sirens - sirens test, fire alarm, danger of flood wave, general warning, chemical accident, radiation accident, ending the alarm. Described is a sequence of processes for the identification of the siren. In Matlab is created by a software application that is being tested on modified warning sirens. The signals are modified noise, offset, gain and attenuation.

Key words

Speaking signal, mean, energy, electronic verbal sirens, verbal sirens recognition.

Seznam použitých symbolů a zkratek

PCM	Pulzní kódová modulace	Pulse Code Modulation
DFT	Diskrétní Fourierova transformace	Discrete Fourier Transform
HMM	Skryté Markovovy modely	Hidden Markov Models
GMM	Gaussian smíšený model	Gaussian Mixture Model
Lram	Délka rámce ve vzorcích	Length frame
Pram	Délka překrytí rámce ve vzorcích	Overlap frame
E	Energie signálu	Signal energy

Obsah

Úvod	1
1. Řeč	2
1.1 Proces vytváření (fyziologie) řeči	2
1.1.1 Ústrojí dechové - respirační	3
1.1.2 Ústrojí hlasové - fonační	3
1.1.3 Ústrojí artikulační	3
1.2 Informační obsah řeči	4
1.2.1 Fonetická forma řeči	4
1.2.2 Akustická forma řeči	4
1.3 Souhlásky v českém jazyce	5
1.3.1 Artikulační vlastnosti souhlásek	5
1.3.2 Akustické vlastnosti souhlásek	7
1.4 Model mechanismu vzniku řeči	8
2. Záznam řeči a její digitalizace	9
2.1 Pulzní kódová modulace	9
2.1.1 Vzorkování	9
2.1.2 Kvantování	10
2.1.3 Kódování	10
3. Předzpracování a analýza akustického signálu	11
3.1 Předzpracování	11
3.1.1 Ustředění	11
3.1.2 Preemfáze	12
3.2 Rozdělení signálu na segmenty	12
3.3 Zpracování v časové oblasti	14
3.3.1 Počet průchodů nulou	14
3.3.2 Střední krátkodobá energie a intenzita	15
3.3.3 Krátkodobá autokorelační funkce	16
3.4 Zpracování ve frekvenční oblasti	17

3.4.1	Krátkodobá diskrétní Fourierová transformace.....	17
3.4.2	Kepstrální analýza.....	18
3.4.3	Výkonová spektrální hustota.....	19
3.5	Určování hraničních bodů promluvy.....	20
4.	Detekce řečové aktivity	21
4.1	Detektory založené na bázi HMM	22
4.1.1	Detektory založené na prahování HMM pravděpodobností.....	22
4.1.2	Kriteriální funkce	22
4.1.3	Algoritmus detektoru využívající segmenty stejné délky	23
4.2	Tradiční detektory	23
4.2.1	Požadavky na detektory	23
4.2.2	Princip stanovení řečové aktivity	23
4.2.3	Typy tradičních detektorů	24
4.3	Tradiční detektory vs. detektory založené na bázi HMM	26
5.	Systém pro detekci verbálního varování a verbální zkoušky sirén	27
5.1	Rozbor průběhu elektronických verbálních sirén.....	28
6.	Návrh systému pro detekci elektronických verbálních sirén	33
6.1	Zobrazení středních hodnot signálů	33
6.1.1	Zobrazení středních hodnot pro 8 rámců.....	36
6.2	Zobrazení energie signálů	38
7.	Realizace systému pro detekci signálu elektronických verbálních sirén.....	46
7.1	Skripty a funkce realizované v programu MATLAB.....	46
7.1.1	Funkce <i>frame.m</i>	46
7.1.2	Skript <i>klasifikace_signalu.m</i>	47
7.2	Uživatelské rozhraní softwarové aplikace.....	49
8.	Vizualizace a srovnání naměřených výsledků	51
8.1	Vliv zesílení a zeslabení klasifikovaného signálu na detekci.....	53
8.2	Vliv rušení klasifikovaného signálu na detekci.....	56
8.3	Vliv posunu klasifikovaného signálu podél časové osy na detekci.....	60
9.	Zhodnocení dosažených výsledků	73
10.	Závěr	76

Literatura.....	77
Seznam příloh.....	79

Úvod

Detekce řečového signálu je významnou součástí mnoha aplikací pro zpracování řeči. Nachází uplatnění v systémech, jak pro zvýraznění řeči, odstranění neřečových částí systémů, přenos řečového signálu, tak pro rozpoznávání a detekci začátku či konce promluvy aj.

První skupinu tvoří algoritmy detekce řeči založené typicky na **výkonové analýze signálu**, **spektrální** či **kepstrální analýze** resp. **koherenční analýze**. Nejjednodušší formy detekce řeči zkoumají **energii signálu** nebo **počet průchodů nulou**. Jejich výhodou je velmi nízká výpočetní náročnost, naopak nevýhodou je vysoká chybovost v případě detekce řeči v šumovém prostředí.

Spolehlivější algoritmy pro detekci jsou založeny na **spektrálních (kepstrálních) vzdálenostech** mezi řečovým signálem a pozadím řeči. V případě zmíněných detektorů se obvykle zjišťuje míra odlišnosti daného bloku signálu od pozadí v dané oblasti (energie, spektrum, entropie spektra). O vlastním výsledku detekce lze pak rozhodnout porovnáním této míry s prahovou hodnotou, kterou lze stanovit globálně jako fixní práh, či ji adaptivně obnovovat podle aktuálních charakteristik pozadí řeči, případně lze využít více sofistikovaných rozhodovacích stromů. Pro telekomunikační systémy se využívají algoritmy, které kombinují několik různých prvků pro zvýšení efektivity detekce.

Další skupinu tedy tvoří algoritmy **detekce na bázi statistického zpracování signálu**. Jedná se o přístup využívající **GMM modelů** řeči a šumu, případně **umělých neuronových sítí** pro nelineární mapování mezi vektorem řečových příznaků a přítomností řeči. Jsou vhodné pro prostředí s velmi vysokou hladinou rušení (např. jedoucí automobil). S jejich nasazením současně vzrůstá výpočetní náročnost detekčních algoritmů a zvyšují se také hardwarové požadavky systému.

V rámci diplomové práce je podrobně rozebrána problematika, jak metod předzpracování, tak samotných systémů pro detekci mluveného signálu.

Dále je zde uveden důkladný popis **návrhu a realizace softwarové aplikace**, pomocí programu Matlab, pro **identifikaci elektronických verbálních sirén** - zkouška sirén, požární poplach, nebezpečí zátopové vlny, všeobecná výstraha, chemická havárie, radiační havárie, konec poplachu.

Úkolem navrženého softwaru je rozpoznání jednotlivých varovných sirén na základě jejich procentuální podobnosti se signály vzorovými. Tato podobnost je vypočtena pomocí energie jednotlivých signálů.

V závěru práce jsou uvedeny **dosažené výsledky**, které byly získány pomocí realizované aplikace. Také jejich vzájemné **srovnání**, včetně **zhodnocení**.

1. Řeč

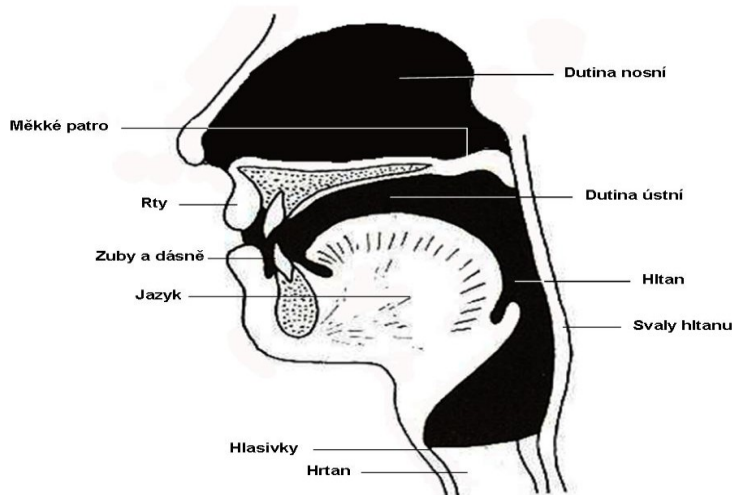
Mluvená řeč je artikulovaný, zvukový **projev člověka sloužící ke vzájemné komunikaci**. Skládá ze slov tvořících lexikální systém určitého jazyka. Přenáší se tzv. komunikačním kanálem a to v podobě akustického signálu. Podstatou tohoto signálu je vlnění elastického prostředí v množině slyšitelných frekvencí. Jsou v něm zakódovány různé druhy informací v podobě akustické složky (amplitudově – frekvenční spektrum) a lingvistické složky, která může mít fonetickou, morfologickou, syntaktickou, sémantickou či pragmatickou strukturu a slouží k vyjádření významu promluvy. Další složka nese spíše specifické informace, které jsou závislé na hlasovém ústrojí a způsobu artikulace, patří zde například intonace, rytmus řeči, barva hlasu apod. Obsahuje také možné anomálie (např. vady řeči) a informace o emocionálním stavu člověka (např. stres, radost). [1]

1.1 Proces vytváření (fyziologie) řeči

Centrum řeči, ze kterého jsou řízeny všechny činnosti související s mluvou, se nachází **v kůře mozkové**. Odtud jsou vysílány koordinované instrukce do svalů, plic, hrtanu, jazyka a obličeje. Orgány zabývající se vytvářením řeči, souhrnně nazýváme řečové orgány (mluvidla) neboli také artikulační orgány (artikulátory). Vytváření řeči nebývá jejich primárním úkolem. Společně se však podílejí na vytváření řeči a seskupení těchto orgánů v těle tvoří hlasový trakt. [1]

Tři základní části hlasového traktu:

- ústrojí dechové - respirační
- ústrojí hlasové - fonační
- ústrojí artikulační



Obr. 1 Vokální trakt a mluvidla u člověka, pohled sagitální (předozadní). [13]

1.1.1 Ústrojí dechové - respirační

Dechové ústrojí představuje pro řeč jakýsi **zdroj energie**. Je tvořeno přírodní cestou dýchací, plicemi a s nimi funkčně spjatými dýchacími svaly (bránicí). Základním materiálem řeči je vzduchový proud výdechu. Zvuk řeči vzniká tak, že se tomuto proudu vzduchu v mluvidlech vytvářejí různé překážky. Trvání výdechu má vliv na to, jak dlouhý úsek řeči lze vytvořit bez přerušení. Vdech znamená pauzu v řeči. [1], [13]

1.1.2 Ústrojí hlasové - fonační

Hlasovým ústrojím se obecně označuje celý **systém pro vytváření řeči** nebo také jen část systému, kde dochází k samostatnému vzniku hlasu. V tom druhém případě mluvíme o aparátu uloženém v hrtanu, který je s plicemi spojen průdušnicí. Z hlediska tvorby řeči, nejdůležitější část hlasového ústrojí tvoří hlasivky. Jsou to dvě ostré slizniční řasy, které vedou napříč hrtanem v místě jeho nejužšího průchodu. Při vytváření hlasu (fonaci) se hlasivky nacházejí v tzv. hlasovém (fonačním) postavení. Výdechový proud vzduchu postupuje bez odporu z plic průdušnicí až k hrtanu. Zde se mu do cesty postaví překážka vytvořená hmotou hlasivek, které cestu vzduchu úplně uzavřou. Stažené hlasivky se pod tlakem vzduchu stávají pružnými a začínají kmitat. V důsledku kmitání hlasivek vzniká základní (hlasivkový) tón, který představuje nosný zvuk řeči. Frekvence kmitání hlasivek se označuje F_0 a nazývá se **frekvence základního hlasivkového tónu**. Tato frekvence nabývá hodnot asi od **60–400 Hz**. U **mužů** se F_0 pohybuje asi mezi **80–160 Hz**, u **žen** je to **150–300** a u **děti** asi **200–400 Hz**. Fonační postavení hlasivek má za následek vznik hlasivkového tónu a používá se proto při vytváření znělých zvuků řeči (samohlásky a znělé souhlásky). Neznělé zvuky jsou naopak tvořeny při klidovém postavení hlasivek, neobsahují tedy základní hlasivkový tón a vznikají tedy až modifikací výdechového proudu vzduchu v nadhrtanových dutinách. [1], [13]

1.1.3 Ústrojí artikulační

Posledním ústrojím je artikulační, i to hraje důležitou roli při tvorbě řeči. Jeho hlavní význam spočívá v tom, že **vytváří velké množství různých zvuků, které charakterizují mluvený jazyk**. Skládá se jednak z nadhrtanových dutin a jednak z artikulačních orgánů, které jsou v těchto dutinách uloženy nebo je obklopují. Mezi nadhrtanové dutiny řadíme dutinu hrdelní, ústní a nosní. Hranici mezi těmito dutinami tvoří čípek (špička měkkého patra), který zamezuje nebo umožňuje přístup vzduchu z dutiny hrdelní do dutiny nosní. Zatímco se nadhrtanové dutiny účastní procesu tvorby řeči pasivně (nepohybují se), artikulační orgány se účastní tvorby řeči většinou aktivně – tvoří pohyblivé součásti artikulačního ústrojí a svým pohybem mění velikost nadhrtanových dutin. Z hlediska vytváření řeči patří mezi nejvýznamnější artikulátory jazyk, rty a měkké patro, neboť se podílejí na vytváření největšího počtu různých zvuků. Dalšími artikulátory jsou zuby, tvrdé patro, ale také hrtan, který svými pohyby může měnit délku celého hlasového traktu. [1], [13]

1.2 Informační obsah řeči

V rámci informačního obsahu řeči se kvantifikuje tzv. informační rychlost (v bitech za sekundu, ve zkratce bit/s nebo bps) nutná k popisu řeči v různých formách. Pro srovnání je uvedena fonetická a akustická forma s číselným vyjádřením signálu.

1.2.1 Fonetická forma řeči

Základní jednotkou fonetických popisů jazyka je **hláska**, v některých koncepcích se mluví o fónech. Jedná se o minimální zvukovou jednotku řeči, zobecněný typ zvukové reality patřící určitému jazyku. Hlázky od sebe odlišujeme podle místa tvoření, podle artikulačního orgánu, nebo také podle sluchového dojmu. Ve světových jazycích se pohybuje počet hlásek od 12 do 60. V českém jazyce jich je **36**, kdežto v anglickém dokonce 42.

Spojování hlásek vytváří posloupnost tzv. **slabik**. Řeč je vlastně pravidelné opakování různých posloupností slabik, jejichž kombinace, vytváří slovo. Slovanské jazyky obsahují **2 500 – 3 500 slabik** a **45 000 – 50 000 slov**.

V rámci standardního rozhovoru člověk vysloví $\pm 80 - 130$ **slovo/min**. Což představuje frekvenci asi **10 fónem/s**. Zdali uvážíme průměrné množství informace na jeden fónem **H = 3 – 4 bit**, dostaneme pro mluvenou řeč **rychlost přenosu informace 30 – 40 bit/s**. Tento poznatek vypovídá o informačním obsahu řeči, který se vyskytuje v její fonetické struktuře.

Z experimentů v rámci psychoakustiky (vnímání zvuku v mozku) bylo zjištěno, že člověk je schopen zpracovat informaci o rychlosti maximálně **50 bit/s**. [2]

1.2.2 Akustická forma řeči

Řeč můžeme zkoumat také z hlediska informačního obsahu skutečného průběhu akustického signálu.

Zvuk obecně můžeme definovat jako mechanické kmitání, které je charakterizováno parametry pohybu částic pružného prostředí nebo u vlnového pohybu parametry zvukového pole. Část zvuků se projevuje jako **slyšitelný zvuk** - což je akustické kmitání pružného prostředí v pásmu frekvencí od **16 Hz do 20 kHz**, schopné vyvolat zvukový vjem. Frekvenční závislost definice slyšitelného zvuku je silně individuální, jen málokdo je schopen vnímat celé pásmo frekvencí (především horní hranice je velmi proměnná a závislá na věku). Zvuky mimo toto pásmo neslyšíme, přesto jsme schopni je vnímat a mohou mít i nepříznivý vliv na zdraví či psychiku. Zvuky pod slyšitelnou hranicí (**0,7 - 16 Hz**) označujeme jako **infrazvuk** (velmi nízké frekvence, lidské tělo je vnímá hmatem - jsou schopny rozvibrovat celý povrch těla či bránici), zvuky nad slyšitelnou hranicí (**do 50 kHz**) jako **ultrazvuk**.

Akustický signál je charakterizován amplitudovým průběhem v čase a také průběhem změn frekvence v čase. Pro kvantitativní vyjádření informačního obsahu zvuku se využívá **Shannonovy věty o výběru**. Tato věta vyjadřuje fakt, že spojitou funkci času, jejíž kmitočtové spektrum je omezené tak, že neobsahuje vyšší kmitočty než F_m , je možno nahradit řadou jejich vzorků, jestliže **frekvence vzorkování** $F_s \geq 2 F_m$.

V případě, že F_m je rovno **12 – 16 kHz** a dynamika řeči **50 dB** při minimálním šumu, je pro dobrý záznam akustického signálu bitová rychlost větší než **200 000 bit/s**. Jedná se o obrovskou informační nadbytečnost. V takové situaci člověk využívá interní mechanismy, které mu umožní potlačit nepotřebné záznamy, jako například intonaci, barvu hlasu apod. a naopak zdůraznit několik hlavních opěrných zvukových příznaků, obsahujících základní informaci, jenž je zakódována v řečovém signálu a shodná pro stejná slova.

Pokud má být výsledkem vnímání řeči identifikace řečníka, bude se posluchač soustředit na jiné charakteristiky mluvy. Problémem stále zůstává fakt, že mechanismy vnímání řeči od zpracování signálu sluchovým orgánem až po porozumění dané zprávě při zpracování v mozku, nejsou v současné době dostatečně známy. Pro nás to znamená, že stále nelze využít jejich analogie při návrhu technických a programových prostředků, jejímž cílem je identifikace řečníka, porozumění řeči apod. Každopádně je vhodné využít, alespoň dostupných znalostí o procesu tvorby řeči a pokusit se odhadnout, které charakteristiky akustického signálu nesou informaci nutnou pro vyřešení konkrétní úlohy. [2]

1.3 Souhlásky v českém jazyce

V českém jazyce se vyskytuje přesně **27 souhlásek**. Při srovnání souhlásek se samohlásky zjistíme, že se souhlásky projevují **charakteristickým šumem a menší amplitudou**, což je stěžním pro srozumitelnost řeči. V důsledku šumu a menší amplitudy je mnohem těžší mezi sebou souhlásky rozlišit, na rozdíl od samohlásek. [1], [5]

1.3.1 Artikulační vlastnosti souhlásek

Artikulace souhlásek je charakteristická tím, že během ní dochází ke **ztížení průchodu vzduchu hlasivkovým ústrojím**. Toto ztížení může být vyvoláno buď to vytvořením překážky, nebo zúžením hlasivkového traktu v daném místě hlasivkového ústrojí. Na základě tohoto faktu vznikají různé zvuky s charakteristickým šumem. Při artikulaci českých souhlásek se využívají čtyři charakteristiky [1], [5]:

- **Způsob artikulace**
- **Místo artikulace**
- **Postavení měkkého patra**
- **Činnost hlasivek**

1.3.1.1 Způsob artikulace

Podle způsobu artikulace dělíme souhlásky na **závěrové** (úplná překážka), **úžinové** (částečná překážka) a **polozávěrové**.

Závěrové souhlásky neboli tzv. **okluzivy**, nazýváme také souhláskami **výbuchovými (exploziv)**. Vytváří se tak, že artikulační orgány uzavřou cestu výdechovému proudu vzduchu. Do této skupiny patří /p/, /t/, /tʰ/, /k/, /b/, /d/, /dʰ/, /g/, /m/, /n/ a /ŋ/.

Úžinové souhlásky – konstriktivy, jsou nazývány také jako **třené souhlásky – frikativy** a vznikají těsným přiblížením artikulačních orgánů, které se však úplně nesevřou. Do skupiny řadíme souhlásky /f/, /v/, /s/, /š/, /z/, /ž/, /j/, /ch/, /h/, /l/, /r/ a /ř/.

- **Sykvky (sibilanty)** – skupina třených souhlásek, při jejichž výslovnosti vznikají výrazné turbulence vzduchu, které se při poslechu projevují intenzivním šumem.

Polozávěrové souhlásky neboli **semiokluzivy** spojují oba typy překážek. V daném místě artikulačního ústrojí se vytvoří krátký závěr, který se však nezruší, ale přejde do úžiny, která se postupně otevře. Souhlásky se nazývají také jako **polootvřené (afrikáty)** a jedná se o /c/, /č/, /dz/ a /dž/. [1], [5]

1.3.1.2 Místo artikulace

Místo artikulace je charakterizováno zúžením nebo úplným uzavřením hlasového traktu. U českých souhlásek se využívá sedm míst hlasivkového traktu.

Dělení souhlásek dle místa artikulace:

- **Retné (labiály)**
 - **obouretné** (/p/, /b/, /m/) - uzavření obou rtů
 - **retozubné** (/f/, /v/) - artikulace dolního retu proti horním řezákům
- **Dásňové (alveoláry)**
 - **předodásňové** (/t/, /d/, /n/, /s/, /z/, /c/, /dz/, /l/, /r/ a /ř/) - špička jazyka artikuluje svou horní plochou proti přední části alveoláry
 - **zadodásňové** (/č/, /dž/, /š/, /ž/) - hřeben jazyka artikuluje proti zadní části alveoláry
- **Patrové**
 - **tvrdopatrové** (/tʰ/, /dʰ/, /ŋ/, /j/) - artikulace střední části hřbetu jazyka proti tvrdému patru
 - **měkkopatrové** (/k/, /g/, /ch/) - artikulace zadní části hřbetu jazyka proti měkkému patru
- **Hrtanové souhlásky (laryngály)** - (/h/) [1], [5]

1.3.1.3 Postavení měkkého patra

Máme i skupinu souhlásek, které využívají pro artikulaci **dutinu nosní**. Zvuk je obohacen o silnou tónovou složku **nosní rezonance**. K tomu dochází, když měkké patro uvolní průchod vzduchu do dutiny nosní. Vznikají tak samohlásky zvané jako **nosové explozivy (nazály)** a patří zde /m/, /n/ a /ň/. [1], [5]

1.3.1.4 Činnost hlasivek

Pokud je foném tvořen základním hlasivkovým tónem - **znělé souhlásky**, vznikají pak další tónové složky zvuku hlásky, způsobené **rezonancí tónu v nadhrtanových dutinách**. Při volném průchodu vzduchu hlasivkami (bez jejich účasti) vzniká zvuk pouze **překážkami v nadhrtanových dutinách - neznělé souhlásky**. Souhlásky tak lze rozdělit do párů, které se liší pouze přítomností základního tónu. Souhlásky nepárové jsou vždy znělé. [1], [5]

1.3.2 Akustické vlastnosti souhlásek

Akustické vlastnosti souhlásek jsou určeny především **podílem tónové a šumové složky**. **Sonorní souhlásky – sonory**, zde převládá tónová složka, proto jsou vždy znělé. Mezi sonory patří nazály /m/, /n/, /ň/, dále pak /l/, /r/ a /j/.

Šumové - pravé souhlásky. Jak už název napovídá, převládá v jejich spektru šumová složka, kromě tónové složky a formantových frekvencí. Řadí se sem všechny párové souhlásky.

Dále lze samohlásky rozdělit podle podobných akustických vlastností:

Likvidy /l/, /r/ a **approximanta** /j/ jsou velmi podobné samohláskám. Jejich časové průběhy jsou také periodické, ale na rozdíl od samohlásek mají menší amplitudu. Jejich spektra jsou si také velmi podobná, u samohlásek bývají o několik decibelů slabší. Šumová složka ve spektru je prakticky nezřetelná.

Nazály, jsou opět svými akustickými vlastnostmi podobné samohláskám. Amplituda signálů nazál je ovšem menší. Na rozdíl od všech hlásek se zde objevují kromě formantů, také antiformanty (spektrální nuly). Ve spektru lze také pozorovat skokové změny amplitudy a frekvence způsoben okluzí. Okluze se ovšem neprojeví pauzou v důsledku postupného průchodu vzduchu nosní dutinou.

Šumové frikativy a /ř/. Akustické signály šumových frikativ jsou spíše neperiodické a mají menší amplitudu než sonory a jejich energie je soustředěna do vyšších frekvencí. Ve spektrech neznělých frikativ se nenacházejí formanty a jejich tvar je podobný horní propusti. Pro znělé frikativy je naopak typická slabá formantová struktura. V případě /ř/ dochází k modulaci šumové složky periodickou frekvencí od 40-60Hz. U všech znělých šumových frikativ a /ř/ předbíhá harmonická složka šumovou o 30-100 ms.

Explozivy, jedná se o přechodové zvuky, kde se okluzní část (závěr) projevuje jako ticho (pauza) s délkou trvání kolem 80 ms. Uvolnění překážky způsobí krátký výbuch šumu (ve spektrogramu podobný vertikálnímu pruhu), který vybudí všechny frekvence. Výbuch přejde ve frikativní šum a následně v artikulovaný foném. Pro explozivы „neuvolněné“ platí, že okluze přechází postupně v akustický signál fonému.

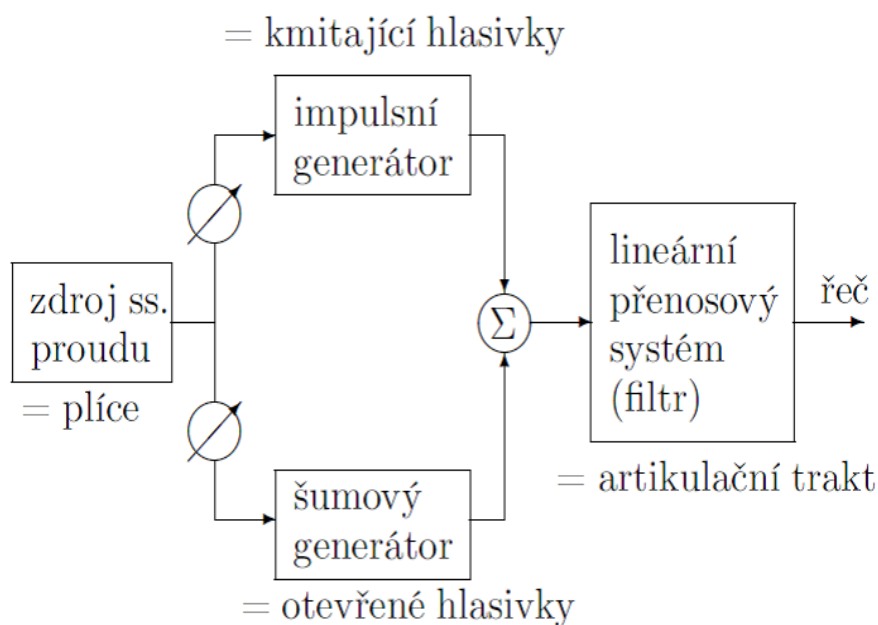
Afrikáty - k jejich vytvoření se využívá všech typů artikulace – okluze, exploze a frikce. Délka trvání afrikáty ovšem není tak dlouhá jako spojení explozivы a frikativы a ani exploze není tak výrazná.

Souhlásky mají jednu důležitou vlastnost a to tu, že jejich délka není tak proměnlivá jako u samohlásek. [1], [5]

1.4 Model mechanismu vzniku řeči

Na základě dosavadních informací lze definovat **blokové schéma mechanismu vzniku řeči** a to pomocí elektronické analogie (řečového syntezátoru).

Zdroj stejnosměrného elektrického proudu představují plíce, které jsou zdrojem proudu vzduchu. Kmitající hlasivky s přítomností F_0 lze považovat za ekvivalent impulsního generátoru a hlasivky otevřené (neznělé) za šumový generátor. Artikulační trakt zde hraje roli lineárního filtru. [4]



Obr. 2 Blokové schéma řečového syntezátoru.

2. Záznam řeči a její digitalizace

Signál můžeme charakterizovat, jako fyzikální veličinu závislou na čase, která má v sobě obsaženou informaci.

Hlavní úlohou lidské řeči je dorozumívání se, tudíž přenášení informace, a proto ji můžeme interpretovat také jako signál. V pružném prostředí se tento signál šíří jako mechanické kmitání. Což ovšem není vhodné pro následné zpracování. **Akustický signál se tudíž pomocí mikrofону převádí na elektrický.** [3]

Z hlediska spojitosti v čase rozlišujeme tři typy signálů:

- Signály analogové (spojité v čase i amplitudě)
- Signály diskrétní (spojité v amplitudě, diskrétní v čase)
- Signály číslicové (diskrétní v čase i v amplitudě)

2.1 Pulzní kódová modulace

Pulzně kódová modulace (PCM z angl. Pulse – code modulation) neboli digitalizace, je modulační metoda **převodu analogového zvukového signálu na signál digitální.**

Akustický signál je zaznamenán mikrofonom. Pro další zpracování je zapotřebí analogové kmity převést do číslicového tvaru a to tak, aby byl spojitý signál reprezentován posloupností číslicových údajů. [2]

PCM zahrnuje tyto kroky:

- Vzorkování
- Kvantizace
- Kódováním

2.1.1 Vzorkování

Vzorkování signálu je proces jeho **diskretizace v časové oblasti**. Znamená to, že vodorovnou osu rozdělíme na rovnoměrné úseky a z každého úseku odebereme jeden vzorek. Vzorky jsou od sebe rovnoměrně vzdáleny se vzorkovací periodou T_S .

Rovnoměrné vzorkování = násobení signálů souvislého času periodickým vzorkovacím signálem. Pro zjednodušení výpočtů zavádíme **ideální vzorkování**, v rámci kterého je vzorkovacím signálem posloupnost Diracových impulsů.

Jednou z důležitých vlastností vzorkovaného signálu je jeho periodicitu spektra. Díky periodicitě vzniká při vzorkování jisté nebezpečí nevratné ztráty informace a to v důsledku

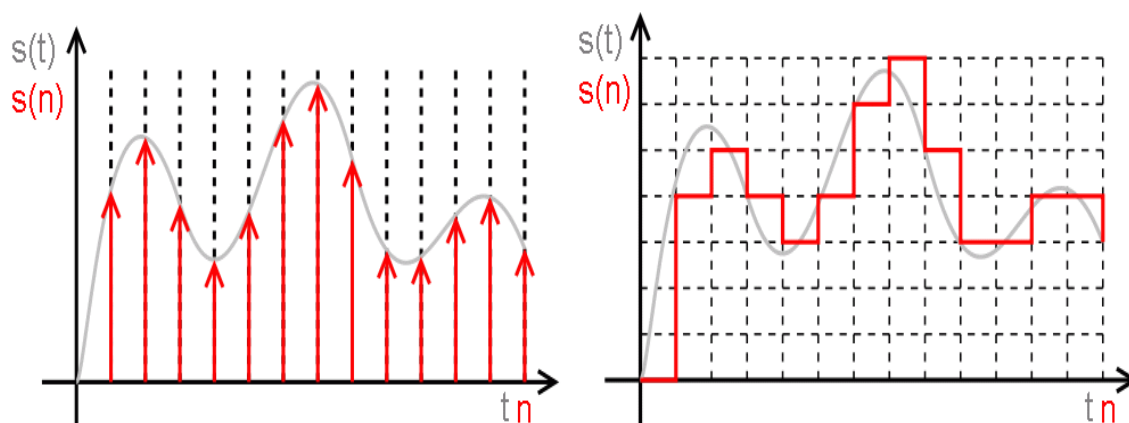
překrytí spekter dvou sousedních period. Prevencí tohoto překrytí, je dodržení Shannonova vzorkovacího teorému. [3]

2.1.2 Kvantování

Kvantování je **diskretizace oboru hodnot signálu**. Jedná se o **proces ztrátový a nevratný**. Pokud není kvantování dostatečně jemné, vznikají falešné kvantizační hrany (kvantizační šum). Nejmenší možný rozdíl mezi dvěma hodnotami kvantovaného signálu se nazývá kvantovací krok. [2]

2.1.3 Kódování

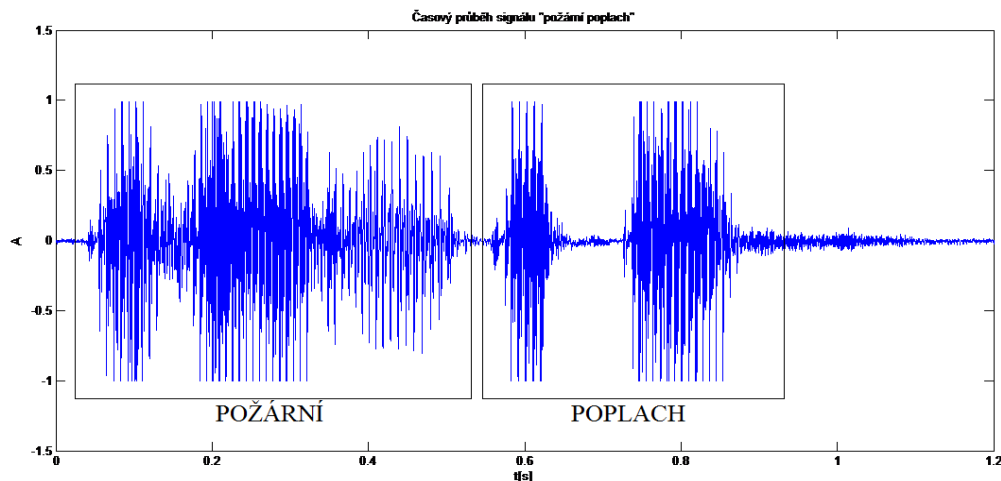
Kvantování s následným kódováním je **přiřazení analogové hodnoty** vzorku signálu jedné hodnotě z konečného počtu číselných hodnot. Každý vzorek je vyjádřen B - bitovým binárním kódem a počet možných úrovní je většinou 2^B . [2]



Obr. 3 Signál se spojitým časem $s(t)$, definován na oboru reálných čísel \mathbb{R} a signál s diskrétním časem $s(n)$, definován na oboru přirozených čísel \mathbb{N} . [14]

3. Předzpracování a analýza akustického signálu

Před samotným rozpoznáním jednotlivých částí řečového signálu, musí být signál náležitě upraven a analyzován. V následující kapitole budou krom předzpracování, rozebrány také jednotlivé možnosti samotné krátkodobé analýzy řečového signálu.



Obr. 4 Časový průběh řečového signálu „požární poplach“.

3.1 Předzpracování

Před samotným zpracováním signálu v časové či frekvenční oblasti je vhodné daný akustický signál předzpracovat. Tím je myšleno **odečtení střední složky** od signálu, která by mohla do budoucna dělat problémy v rámci jiných výpočtů a aplikace **preemfáze** neboli zvýraznění vyšších kmitočtů.

3.1.1 Ustředění

Na začátku zpracování je důležité **odečíst od signálu jeho střední hodnotu**. Stejnosměrná složka nenese žádnou potřebnou informaci. Spíše může být v rámci některých výpočtů škodná (například výpočet krátkodobé energie).

Co se týče výpočtu střední hodnoty, je složitější, pokud neznáme průběh celého signálu. [4]

V případě, že signál známe [4]:

$$\bar{s} = \frac{1}{N} \sum_{n=1}^N s[n] \quad 3.1$$

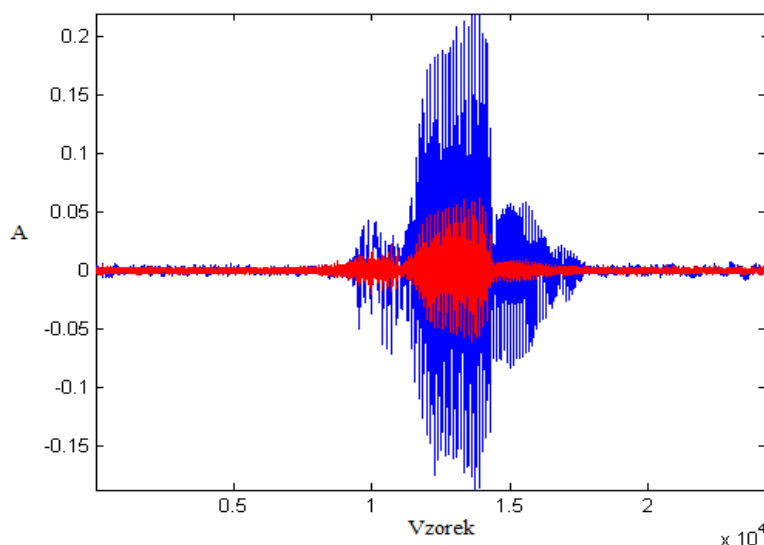
Ncelkový počet vzorků signálů

$s[n]$...jednotlivé vzorky řečového signálu

3.1.2 Preemfáze

Preemfáze, jednoduše řečeno, slouží ke **zvýraznění vyšších kmitočtů**. Přenos vyšších kmitočtů je zdůrazněn oproti přenosu nízkých kmitočtů a to podle daného předpisu. Opakem preemfáze je tzv. deemfáze, díky které se zvýrazněné vyšší kmitočty sníží na původní úroveň.

V rámci zpracování akustického signálu **vyrovnává kmitočtové charakteristiky řeči**, což znamená, že energie řeči klesá směrem k vyšším frekvencím. Jedná se spíše o historickou operaci. [4], [15]

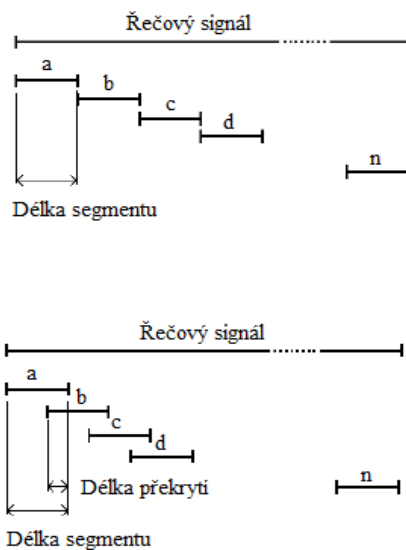


Obr. 5 Originální signál – modře, preemfáze signálu – červeně. [15]

3.2 Rozdělení signálu na segmenty

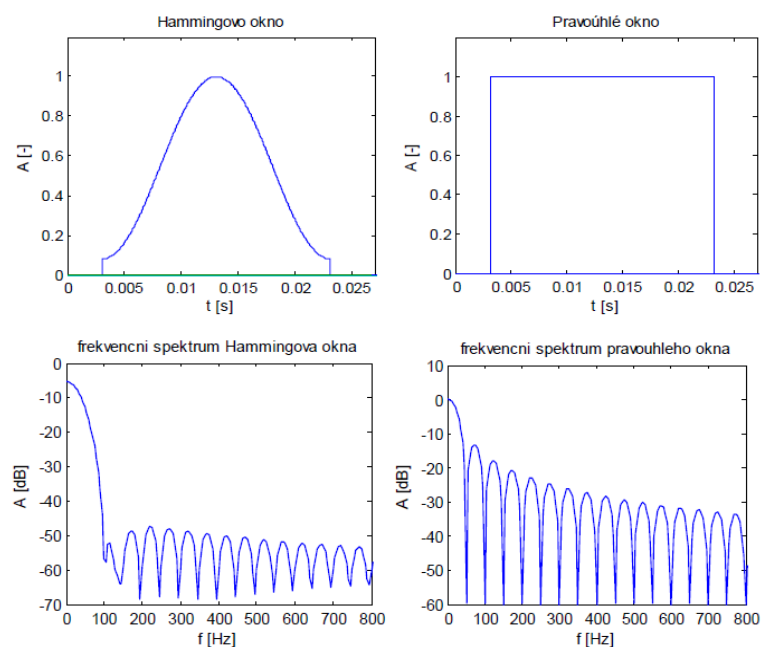
Akustický signál je nutno před dalším zpracováním rozdělit na tzv. **rámce**. Tento krok se provádí, neboť metody sloužící pro odhad parametrů příznivě pracují se **stacionárním signálem**. Stacionární signál má oproti nestacionárnímu signálu v čase neměnné statické vlastnosti (střední hodnotu, distribuční funkci, rozptyl atd.). V případě, že chceme zpracovávat stacionární úsek, musíme brát v potaz setrvačnost hlasového ústrojí. Z tohoto důvodu je signál dělen většinou na úseky dlouhé **20 – 25 ms**. Délka rámce ve vzorcích při $F_S = 16 \text{ kHz}$ je $L_{ram} = 320 – 400$ vzorků.

Je vhodné, aby se jeden úsek řeči stal součástí několika rámců. Jinými slovy, aby se jednotlivé sousední rámce z určité části překrývaly. Problém může nastat s rostoucí hodnotou překrytí, kdy se také zvýší nároky na paměť a podobnost sousedních rámců. Na druhou stranu s malým překrytím se hodnoty parametrů mezi sousedními rámci mohou hodně měnit. Proto je nejlepším řešením kompromis, kterým je většinou překrytí $P_{ram} = 10 – 15 \text{ ms}$.



Obr. 6 Princip segmentace bez překrytí sousedních segmentů a s překrytím sousedních segmentů.

Proces rámcování signálu se uskuteční pomocí vykrojení tzv. okna. Nejčastěji se používá **Hammingovo okno** nebo **okno trojúhelníkového tvaru**. Pravoúhlé okno má ve frekvenčním spektru užší hlavní svazek a výrazné postranní laloky, kdežto Hammingovo okno má širší hlavní svazek a postranní laloky značně tlumeny. [4]



Obr. 7 Srovnání frekvenční a časové charakteristiky pravoúhlého a Hammingova okna.[4]

Pro rozdělení signálu na segmenty byl v **MATLABu** vytvořen skript *frame.m*. Jeho vstupními parametry jsou:

- **Zvukový signál, zadaný jako sloupcový vektor**
- **Vzorkovací kmitočet signálu**
- **Délka segmentů zadána v sekundách**
- **Délka překrytí segmentů zadána v sekundách**

Výstupem je matice, jejíž sloupce jsou jednotlivé segmenty. Index sloupce odpovídá pořadí segmentu a index řádku pořadí vzorku v segmentu. Takto rozdělený signál je připraven k dalšímu zpracování.

3.3 Zpracování v časové oblasti

Při zpracování v časové oblasti se vychází přímo z **hodnot vzorků signálu**. Příznaky získané z časové oblasti jsou většinou méně náročné na výpočet. Mezi metody využívající zpracování v časové oblasti patří například:

- **Krátkodobá střední hodnota průchodu signálu nulovou úrovní**
- **Krátkodobá energie**
- **Krátkodobá intenzita**
- **Krátkodobá autokorelační funkce**

3.3.1 Počet průchodů nulou

Po rozdělení signálu na rámce, se pro každý z nich stanoví **počet průchodů nulou Z**. Jedná se o parametr, který **slouží k určení znělých a neznělých hlásek**.

Už víme, že pro tvorbu neznělých hlásek se nepoužívají hlasivky, proto se více podobají šumu a mají vyšší Z. Někdy je problém rozlišit neznělou hlásku od šumu a to z důvodu, že počet průchodů nulou je na šum značně náchylný.

Pro získání tohoto počtu se v **MATLABu** využívá **funkce sign**.

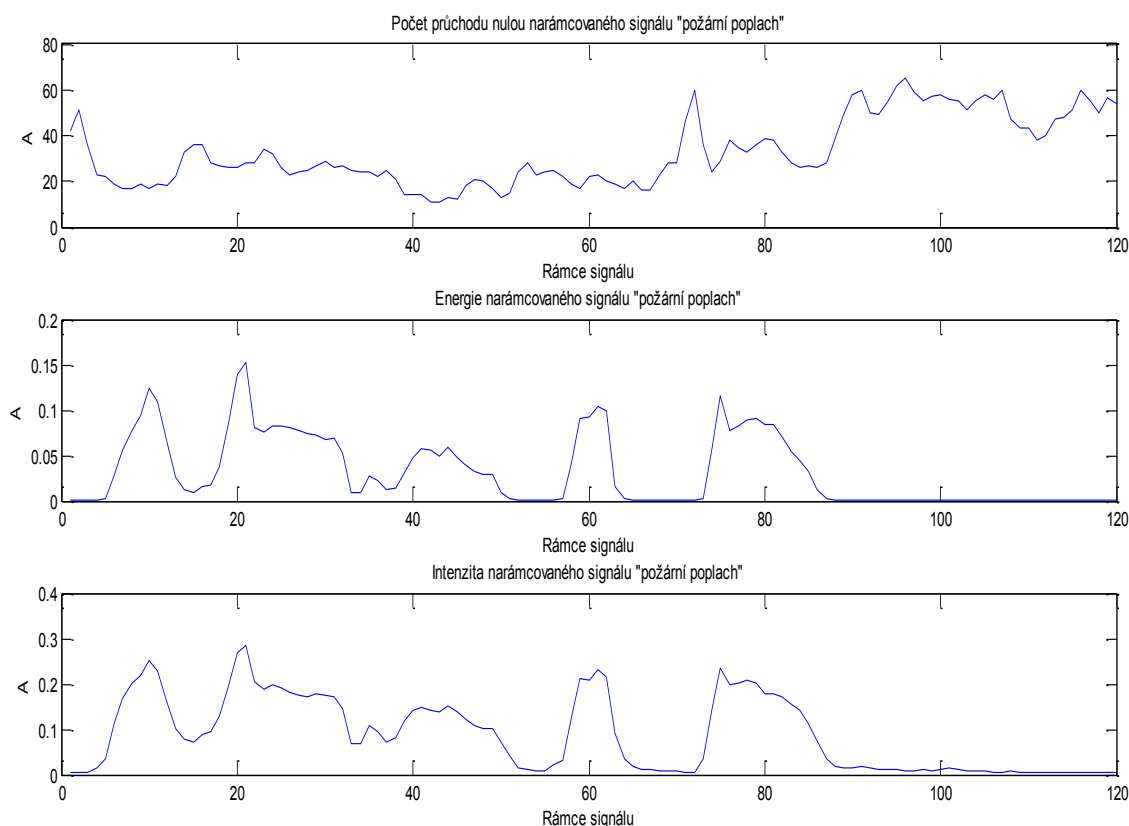
Vztah pro výpočet průchodů nulou Z [2]:

$$Z = \frac{1}{2} \sum_{n=1}^{L_{ram}} | \text{sign}x[n] - \text{sign}x[n-1] | \quad 3.2$$

$$\text{kde } \begin{cases} \text{sign}x[n] = 1 & \text{pro } x[n] > 0 \\ \text{sign}x[n] = -1 & \text{pro } x[n] < 0 \\ \text{sign}x[n] = 0 & \text{pro } x[n] = 0 \end{cases}$$

Při průchodu signálu nulou je hodnota v sumě rovna dvěma. Proto se výsledek musí o polovinu zmenšit.

V případě, že signál začíná v nule, je hodnota v sumě rovna 1 a ve výsledku se to projeví hodnotou o 0,5 větší, což je problém. [2]



Obr. 8 Počet průchodů nulou, energie a intenzita vzorového signálu „požární poplach“.

3.3.2 Střední krátkodobá energie a intenzita

Také střední krátkodobá energie může být využita pro **rozlišení znělých a neznělých hlásek**. Znělé hlásky jsou charakteristické vyšší energií ve srovnání s neznělými. Tato energie se také využívá pro **detekci řečové aktivity**. Jedná se o nejjednodušší způsob, jakým detekovat řeč. I tato metoda je více než dost náchylná na šum, zejména u neznělých hlásek často selhává. [2]

Vztah pro krátkodobou energii E [2]:

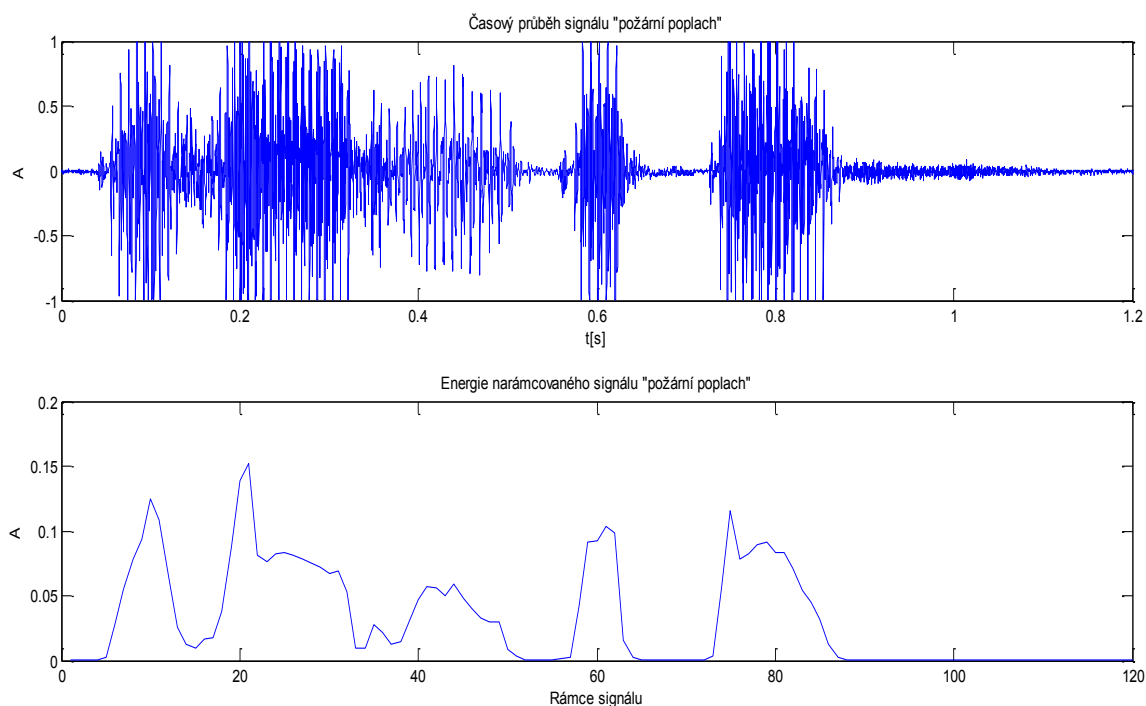
$$E = \frac{1}{L_{ram}} \sum_{n=0}^{L_{ram}-1} x[n]^2 \quad 3.3$$

Pro výpočet krátkodobé energie byl v **MATLABu** vytvořen skript **energie.m**. Jeho vstupní hodnotou je segmentovaný řečový signál a výstupem řádkový vektor příznaku energie pro každý mikrosegment.

Pro zmenšení dynamického rozsahu se často energie uvádí v decibelech nebo se pracuje s krátkodobou intenzitou.

Vztah pro krátkodobou intenzitu M [2]:

$$M = \frac{1}{L_{ram}} \sum_{n=0}^{L_{ram}-1} |x[n]| \quad 3.4$$



Obr. 9 Časový průběh signálu „požární poplach“ a krátkodobá energie tohoto signálu.

3.3.3 Krátkodobá autokorelační funkce

Autokorelační funkce má některé vlastnosti, díky kterým může být použita pro určení periodicity signálu. Autokorelační funkce v podstatě udává **podobnost signálu na sebe samého**, posunutého o n vzorků. Pro výpočet autokorelační funkce by měl mikrosegment obsahovat alespoň dvě periody signálu, musí být tedy dostatečně dlouhý (udává se 20 - 40 ms).

Z metod zpracovávající signál v časové oblasti má autokorelační funkce největší výpočetní nároky, za to je v některých případech efektivnější. [1]

Vztah pro krátkodobou autokorelaci R [1]:

$$R(m) = \frac{1}{L_{ram}} \sum_{n=0}^{L_{ram}-1} x(n) x(n + m) \quad 3.5$$

Autokorelační funkce je poslední z metod zpracovávající signál v časové oblasti využitých při segmentaci řečového signálu na fonémy.

Postup výpočtu autokorelace v **MATLABu**: vstupem je opět segmentovaný řečový signál a výstupem zde pro změnu není skalární hodnota pro každý segment, jak tomu bylo u předchozích metod, ale vektory koeficientů. Pro výpočet autokorelace se používá **funkce xcorr**. Pro dva vstupní signály tato funkce určí jejich **vzájemnou korelaci** a při jednom vstupním signálu jeho **autokorelaci**.

3.4 Zpracování ve frekvenční oblasti

Mluvená řeč je ve frekvenční oblasti reprezentována svým spektrem, velikostí frekvenčních složek. Základem většiny metod zpracování ve frekvenční oblasti je **diskrétní Fourierova transformace**.

Výpočetní náročnost metod zpracování ve frekvenční oblasti je často větší než u metod z časové oblasti. Většinou však oproti nim dávají přesnější výsledky. Jsme schopni detekovat hranice, které by metodami z časové oblasti zůstaly nenalezeny.

3.4.1 Krátkodobá diskrétní Fourierová transformace

Abychom **oddělili nepotřebné informace od signálu**, je důležité jej převést do frekvenční oblasti. Využívá se Diskrétní Fourierovy transformace (DFT).

DFT je dána vztahem [2]:

$$X(k) = \sum_{n=0}^{N-1} x[n] e^{-j2\pi \frac{nk}{N}}, \text{ pro } k \in \langle 0, N-1 \rangle \quad 3.6$$

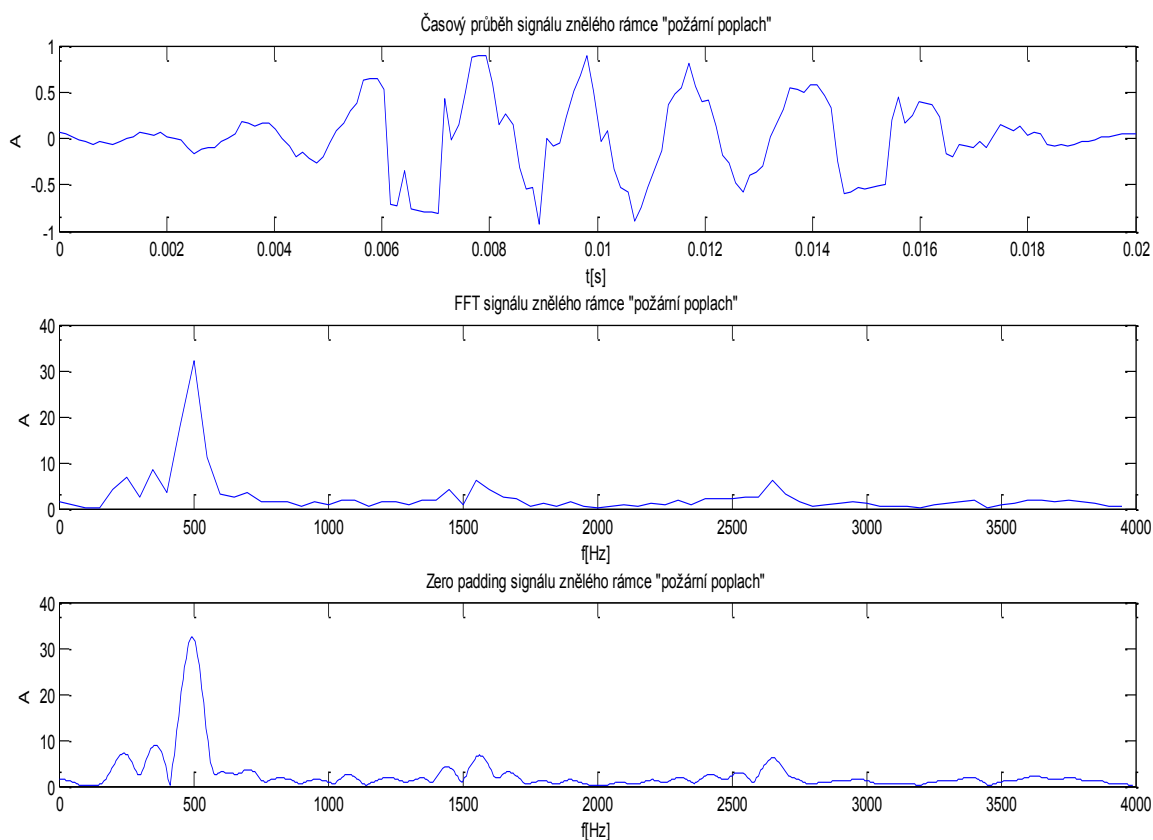
Tímto vztahem se vypočítá spektrum signálu, jenž je periodické a to s periodou rovnou vzorkovací frekvenci. V případě, že se bude DFT aplikovat na jednotlivé rámce, projeví se ve spektru i to, jaké bylo použito okno pro tvorbu rámce. Hodnota spektra je diskrétní s hodnotami kmitočtů vzdálených $f_v = F_s / N$. Výsledné spektrum je symetrické, a proto stačí vykreslit jen jeho první polovinu. [2]

Pro výpočet diskrétní Fourierovy transformace byl v **MATLABu** vytvořen skript **dft.m**. Základem tohoto skriptu je algoritmus FFT, který je realizován pomocí **funkce fft**. Po aplikaci této funkce na vstupní segmentovaný signál získáme komplexní spektrum každého mikrosegmentu. Z něj se vypočte modulové spektrum jako absolutní hodnota komplexního spektra a to pomocí **funkce abs**. Spektrum reálného signálu získané pomocí **fft** je symetrické kolem poloviny vzorkovacího kmitočtu, a proto pro další výpočty stačí použít jen první polovinu vypočteného modulového spektra.

3.4.1.1 Zero padding

Už jsme seznámeni se vztahem, který definuje vzdálenost jednotlivých vzorků spektra. Mohou však nastat případy, kdy je zapotřebí přesnější charakteristika, tudíž více vzorků ve spektru.

Výsledků se docílí použitím metody „zero padding“ neboli metody doplňování nul. Doplněním nul se zvýší N , ale žádná informace se do výpočtu spektra nepřidá. [4]



Obr. 10 Frekvenční charakteristika znělého rámce signálu „požární poplach“ bez a s využitím „Zero padding“.

3.4.2 Kepstrální analýza

Kepstrální analýza je metoda umožňující ze signálu řeči **oddělit parametry buzení a hlasového ústrojí**. Každá složka spektra řeči je dána konvolucí buzení a impulzní odezvou hlasového ústrojí. Pro analýzu řeči je vhodné tyto dva parametry oddělit. Ze signálu získaného konvolucí dvou vstupních signálů, je ale obtížné dostat tyto signály zpět. Lze se o to pokusit tak, že zavedeme vhodnou nelineární operaci, například logaritmování, která převede součin na součet. Jednotlivé složky součtu pak od sebe oddělit lze.

Proces kepstrální analýzy se skládá z následujících kroků:

- **Diskrétní Fourierova transformace**
- **Logaritmus spekter mikrosegmentů**
- **Zpětná Fourierova transformace**

Postup výpočtu kepstra v **MATLABu**: pro tento výpočet je v **MATLABu** definovaná **funkce rceps**. Uvnitř **funkce rceps** je implementován algoritmus $y = \text{real}(\text{ifft}(\log(\text{fft}(x))))$.

Pro hledání hranic fonémů metodou sledování příznaků není vhodné použít celé získané kepstrum, ale jen několik prvních koeficientů. [4]

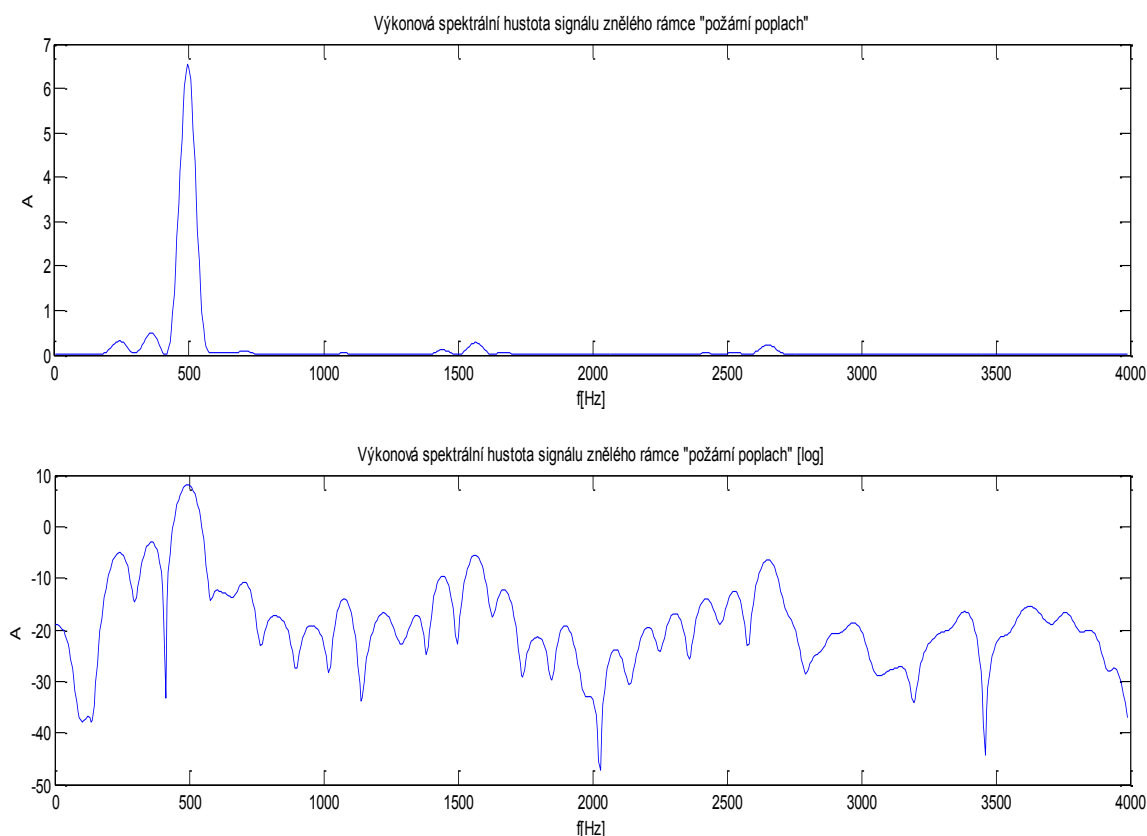
3.4.3 Výkonová spektrální hustota

Výkonová spektrální hustota G se využívá k analýze náhodného signálu a udává **rozdělení výkonu ve frekvenční oblasti**. Jeden z odhadů výkonové spektrální hustoty využívá Diskrétní Fourierovy transformace [4]:

$$G_{DFT}(k\Delta f) = \frac{1}{N} |[X]_k|^2 \quad 3.7$$

Na charakteristice výkonové spektrální hustoty nelze z důvodů vysoké dynamiky pozorovat méně výrazné části, a proto se výsledek logaritmuje a uvádí v decibelech [4]:

$$G_{DFT}(k\Delta f) = 10 \log_{10} \left(\frac{1}{N} |[X]_k|^2 \right) \quad 3.8$$



Obr. 11 Výkonová spektrální hustota signálu znělého rámce a její logaritmus.

3.5 Určování hraničních bodů promluvy

Problémem při zpracování řečového signálu může být určení hraničních bodů. Hraničními body myslíme **začátek a konec promluvy**. Analyzátory většinou nepřetržitě zpracovávají veškerý vstupní signál snímaný z mikrofону a jeho reprezentativní parametry dále předávají klasifikátoru.

V případě, že je zabezpečen vysoký odstup signálu od šumu, což znamená tichou místnost, kvalitní mikrofón apod., problém nalezení hraničních bodů signálu lze vyřešit snadno a to na základě **zvýšené úrovně intenzity signálu řeči**. Tato intenzita i pro nejslabší zvuky, jako jsou neznělé sykavky, převyšuje intenzitu okolního šumu. Jedná se o případy, které nejsou běžné, v reálných aplikacích pracují především klasifikátory slov v lokalitách s podstatným rušivým pozadím.

Pro určení hraničních bodů, je rozhodující průběh intenzity. Jsou však případy, kdy pro přesné stanovení nestačí. A proto je pro korektní vymezení hraničních bodů vhodné využít i **spektrálních charakteristik řečového signálu**. Jedna z metod využívá **průběhy krátkodobé intenzity M** a krátkodobé funkce **středního počtu průchodů signálu nulou Z** . [2]

4. Detekce řečové aktivity

Detekce řečové aktivity hraje důležitou roli v oblasti **zpracování řeči** a je objektem současného výzkumu. Rozpoznat libovolné slovo z mluvy neznámého řečníka se stále dokonale nedaří. Hlavní důvody se skrývají v obrovské variabilitě mluvčího, v prostředí, ve kterém se záznam provádí, ale také v obtížnosti řešené úlohy.

Každý jedinec má originální hlasové ústrojí a odlišný způsob artikulace, což se projevuje odlišnou barvou hlasu, rychlostí řeči, přízvukem apod. Samotný hlas řečníka je velmi variabilní a mimo jiné také závislý na mnoha aspektech, jako je otázka, příkaz, nálada atd. Tyto aspekty se projevují v délce jednotlivých úseků řeči a v intenzitě řečového signálu. Je tedy nepravděpodobné, aby jedno slovo bylo dvakrát řečeno stejně.

Rozpoznávání řeči, dle rozsahu slovníku, lze rozdělit do tří základních skupin:

- **Rozpoznávání izolovaných slov** (malý slovník, např. číslovky, povely)
- **Rozpoznávání diskrétního diktátu** (rozsáhlejší slovník, slova jsou vyslovována izolovaně s krátkou mezislovní pauzou)
- **Rozpoznávání souvislé řeči** (slovník na desítky tisíc slov)

Co se týče samotných detektorů pro rozpoznávání řeči, k dispozici máme **tzv. tradiční detektory**, které řeší tuto úlohu pomocí klasifikátorů založených na **prahování vhodných řečových charakteristik**. Detektory využívající prahování rozhodují o absenci nebo přítomnosti řeči na základě srovnávání hodnot parametrů segmentů s prahovou hodnotou, tj. prahem.

Principiálně můžeme rozlišit **dva základní druhy prahování** – první způsob spočívá ve využití **statického prahu** (detektory na bázi HMM) a druhý způsob využívá **práh dynamický**. Mezi nejznámější typy pracující na tomto principu patří **detektory keprální a energetické**. [7], [8]

Základní detektory řečové aktivity:

- **Statické nastavení prahu**
 - Detektory řeči na bázi HMM (skryté Markovovy modely)
 - Detektory řeči na bázi GMM (Gaussian Mixture Model)
 - Detektory řeči na bázi neuronové sítě (ANN, MLP)
- **Dynamické nastavení prahu (tradiční detektory)**
 - Ideální detektor
 - Energetické detektory
 - Keprální detektory

4.1 Detektory založené na bázi HMM

Tento typ detektoru řečové aktivity je založen na **skrytých Markovových modelech**, které jsou statickými stavovými automaty. Dají se použít jak k rozpoznávání řeči, tak k samotné detekci řeči. Modelují prvky řeči pomocí stavů a přechodů mezi nimi. Stavy modelů jsou charakterizovány pravděpodobnostním rozdělením parametrů určité části signálu, která je daným stavem modelována. Částmi signálu rozumíme například fonémy, slabiky nebo slova. Ergodické Markovovské modely jsou specifické tím, že umožňují přechody mezi libovolnými stavy. Tato struktura je vhodná pro modelování nestacionárních šumových složek signálu.

Detektory tohoto typu **dosahují lepších výsledků než tradiční** a to hlavně u silně rušených signálů. [8]

4.1.1 Detektory založené na prahování HMM pravděpodobností

Detekce probíhá na základě **modelování částí řeči pomocí jednoho nebo i více HMM**. Jeden model většinou odpovídá úsekům řeči a druhý úsekům šumového pozadí. Bude použit pojem **dlouhodobý a krátkodobý segment**.

Z krátkodobého segmentu řeči je vypočítán jeden vektor parametrů. Pojmem dlouhodobý segment se myslí posloupnost vektorů parametrů. V dlouhodobých segmentech stejné nebo rostoucí délky se provádí výpočet pravděpodobnosti, zda HMM generuje příslušný úsek šumu nebo řeči. Výstupem jsou vektory pravděpodobností příslušné ke každému řečovému rámci. Je-li použit pouze jeden model – model šumu, může být prahována přímo tato pravděpodobnost. Jsou-li použity dva modely – model šumu i model řeči, jsou pravděpodobnosti nejprve vzájemně odečteny a výsledný rozdíl je prahován. Tyto vektory, které jsou prahovány, jsou označeny jako kritériální funkce. [8]

4.1.2 Kritériální funkce

Je-li použit jeden **model šumu**, který se označí λ_N , potom je **kritériální funkce $f[n]$** tvaru [8]:

$$f[n] = \log (P(O|\lambda_N)) \quad 4.1$$

Jsou-li použity dva modely: **model šumu λ_N** a **model řeči λ_S** , potom má **kritériální funkce $f[n]$** tvar [8]:

$$f[n] = \log (P(O|\lambda_S)) - \log (P(O|\lambda_N)) \quad 4.2$$

O...dlouhodobý segment řeči

4.1.3 Algoritmus detektoru využívající segmenty stejné délky

U tohoto typu detektoru dochází k analýze HMM pravděpodobnosti (získané z jednoho nebo dvou modelů) z posunovaných **dlouhodobých segmentů stejné délky**.

Algoritmus detekce má 4 fáze [8]:

- **Inicializační fáze:** Zde se nastavují parametry modelů řeči λ_S a šumu λ_N
- **Trénování:** Modely se trénují pomocí Baum-Welchova algoritmu
- **Kriteriální funkce:** V této fázi se vypočte kriteriální funkce
- **Fáze klasifikace:** V poslední fázi dochází k prahování, detekci a vyhlazení mediánovým filtrem

4.2 Tradiční detektory

Tradiční detektory řečové aktivity pracují na základě algoritmu, který následně **klasifikuje čistý řečový signál, či směs řeči a šumu do dvou tříd** a to třídy řeči a třídy šumu.

Řečový signál je rozdělen do segmentů a ke každému segmentu je přidělena příslušná třída. Často se můžeme setkat s označením řeči a šumu jako „1“ a „0“. Výstupem je „1“, je-li v i-tém segmentu přítomná řeč a „0“ je-li segment bez řečové aktivity. Detekci řečové aktivity můžeme provádět i na úrovni vzorků řečového signálu. Průběh detekce řeči je téměř vždy rozdělen do dvou částí. V první jsou ze segmentů řečového signálu získány parametry řeči, které zdůrazňují rozdíl mezi řečí a šumem. Ve druhé fázi probíhá samotná klasifikace. [10]

4.2.1 Požadavky na detektory

Detektory řečové aktivity by měly splňovat předem stanovené **požadavky**, které jsou uvedeny níže [10]:

- Detektor by měl být zrealizovaný ve frekvenční oblasti
- Informaci o přítomnosti řeči by měl získat pouze v krátkých aktuálních signálových segmentech a pracovat v reálném čase
- Detektor by měl dosahovat velké pravděpodobnosti správné detekce řeči, pauzy i při relativně malém poměru signál/šum

4.2.2 Princip stanovení řečové aktivity

Většina detektoru využívá obdobný **princip stanovení řečové aktivity**. Ten lze shrnout do několika bodů [10]:

- Vstupní signál se rozdělí na časové segmenty (rámce)
- Stanovení potřebné charakteristiky signálu (dle detektoru např. energie, kepsrum)

- Vypočtená charakteristika se v každém rámci porovná s prahovou hodnotou
- Prahová hodnota může být v průběhu detekce přepočítávána a aktualizována (detektory s adaptivně nastavovaným prahem), nebo též stanovena pevně
- Je-li dána charakteristika signálu v rámci větší než prahová, je detekována řeč, jinak pauza

4.2.3 Typy tradičních detektorů

Skupina tradičních detektorů zahrnuje **ideální detektor**, realizovaný pomocí keprstrálního šumu na řečový signál bez šumového pozadí. **Energetický detektor**, jehož princip spočívá ve výpočtu energie signálu a **keprstrální integrální detektor**.

4.2.3.1 Ideální detektor

Realizace ideálního detektoru probíhá tak, že se ručně označí úseky signálu s řečovou aktivitou. Jedná se však o velmi pracnou detekci, která se nedá použít pro větší množství dat a **nelze s ní ani detekovat signál v reálném čase**. V praxi se realizuje pomocí aplikací keprstrálního šumu na řečový signál bez šumového pozadí. [11]

4.2.3.2 Energetický detektor

Princip energetického detektoru spočívá ve **výpočtu energie signálu** pro každý rámec. Využívá se zde funkce, která je definována takto:

$$E_n = \sum_{k=-\infty}^{\infty} [s(k)\omega(n-k)]^2 \quad 4.3$$

$s(k)$vzorek signálu v čase k

$\omega(n)$...příslušný typ váhovacího okénka

Při měření **krátkodobé energie** je doporučená délka rámců **10 až 20 ms** při frekvenci vzorkování **8–10 kHz**. Funkce je citlivá na velké změny úrovně signálu, proto se často využívá **krátkodobá intenzita**, která tento nedostatek nemá:

$$M_n = \sum_{k=-\infty}^{\infty} |s(k)|\omega(n-k) \quad 4.4$$

Takto vypočtený segment se porovnává s **prahovou hodnotou energie E_p** , která je definovaná jako:

$$E_p = 1,5E_d \quad 4.5$$

E_d je **úroveň energie hluku pozadí** aktualizována podle vztahu:

$$E_d = (1 - p)E_d^0 + pE \quad 4.6$$

E_d^0 ...poslední známá hodnota energie

Eenergie signálu v segmentu

pparametr (se volí např. 0,5)

Aktualizace se provádí v řečových pauzách. Je-li $E_n > E_p$ je detekována řeč, v opačném případě je energie šumu aktualizována. [10]

4.2.3.3 Kepstrální integrální detektor

Kepstrální detektory vykazují **zvýšenou robustnost vůči šumovému pozadí řeči** v porovnání s jednoduššími algoritmy, např. energetickými.

Vstupní signál značíme $x_l[n]$ (index „l“ znamená l-tý rámec vstupního signálu). Používá-li se při výpočtu reálná část, hovoří se o **reálném kepstru**:

$$c_l[k] = \text{Re}\{IDFT\{\log|DFT\{x_l[n]\}|\}\} \quad 4.7$$

Vybraná charakteristika signálu $x_l[n]$ se u kepstrálního integrálního detektoru řeč/pauza stanovuje z **kepstra signálu v l-tém rámci** $c_l[k]$ a **průměrného kepstra pozadí** $\bar{c}_l[k]$. Hovoří se o tzv. **vzdálenosti kepstra signálu v rámci**, značíme Δc_l :

$$\Delta c_l = 4,3429 \sqrt{(c_l[0] - \bar{c}_l[0])^2 + 2 \sum_{k=1}^p (c_l[k] - \bar{c}_l[k])^2} \quad 4.8$$

Používá se **reálné kepstrum signálu** $x_l[n]$ v aktuálním rámci l o délce N vzorků. Detekce řečového úseku se provádí porovnáním **kepstrální vzdálenosti** Δc_l :

1 – pokud je $\Delta c_l \geq t_l$

0 – pokud je $\Delta c_l < t_l$

V řečové pauze se aktualizuje **prahová hodnota** t_l . Využívá se matematické statistiky:

$$t_l = E(\Delta c[n]) + \alpha \sqrt{D(\Delta c[n])} \quad 4.9$$

Také dochází k aktualizaci **průměrného kepstra pozadí** $\bar{c}_{l+1}[k]$:

$$\bar{c}_{l+1}[k] = (1 - \beta) \bar{c}_l[k] + \beta c_l[k] \quad 4.10$$

Poznámky:

α – je vhodné volit v intervalu 1,8 až 3,5,

β – časová konstanta exponenciálního průměrování (nabývá hodnot 0 až 1).

Tento typ detektoru potřebuje pro správnou činnost tzv. **počáteční prodlevu** neboli inicializační fázi, během které je z několika počátečních segmentů spočtena **střední hodnota** a **rozptyl keprální vzdálenosti**. [7], [9]

4.3 Tradiční detektory vs. detektory založené na bázi HMM

V případě použití statického výpočtu prahu, je nevýhoda tohoto způsobu detekce řeči přímočará, při významnější změně prostředí začíná takto postavený detektor selhávat. Tento problém se snaží řešit druhý přístup využívající dynamické nastavení prahu. Výhodou je, že jsou tyto algoritmy ustálené a dlouhodobě spolehlivé. Další výhodou tohoto postupu je, že uživatel má plnou kontrolu nad rozhodováním, protože rozhoduje nad pravidly, která slouží k výpočtu hodnoty prahu. Naopak nevýhodou je, že nacházení prahu může být velmi složité, a to zvláště v případech, kdy je k dispozici rozsáhlá množina parametrů signálu. Možnosti zlepšení algoritmů detekce řeči je v hledání nových postupů pro získávání aktuální prahové hodnoty nebo využití jiných algoritmů, které prahování vůbec nepoužívají. Další výhodou by bylo využití algoritmů, které jsou schopny nastavit parametry detektorů řeči tak, že jsou z určitého hlediska optimální. Tyto body splňuje využití statických klasifikačních algoritmů, mezi které například patří skryté Markovovy modely (HMM). [8]

5. Systém pro detekci verbálního varování a verbální zkoušky sirén

Cílem diplomové práce je tvorba softwarového programu v prostředí **MATLAB**, který je schopen klasifikovat **verbální varování a verbální zkoušky sirén**, sdělované obyvatelstvu. Tyto údaje mají povahu hrozby nebo již nastalé mimořádné události (např. chemická havárie, radiační havárie, nebezpečí zátopové vlny apod.) a opatření k ochraně obyvatelstva.

„Rozpoznávání verbálních varování a zkoušek sirén může být ideálním řešením například pro neslyšící, kterým bude blížíci se hrozba sdělována například pomocí specifických vibrací, obrázku na displeji mobilního zařízení apod.“

Zkouška nebo skutečné varování prostřednictvím sirén jsou od sebe jasně odlišitelné. **Zkouška sirén** probíhá zpravidla **nepřerušovaným tónem** po dobu **140 s**.

K **varování** obyvatelstva před hrozbou nebo při vzniku mimořádné události je používán jeden varovný signál pod názvem „**Všeobecná výstraha**“. Tón je **kolísavý**, trvá **140 s** a může zaznít v cca **tříminutových intervalech**.

Sirény kromě všeobecné výstrahy a varování slouží také k **přivolávání jednotek požární ochrany**. Tento signál, je vyhlašován charakteristickým **přerušovaným tónem** sirény po dobu **jedné minuty**.

Na celém území České republiky je nainstalováno, prověřováno a udržováno v provozuschopném technickém stavu více jak 5 000 sirén, z nichž většinu je možné ovládat dálkově. Tyto sirény pokrývají 85% obydleného území ČR varovným signálem.

Typy sirén [12]:

- **Elektronické verbální** - nejmodernější, jsou opatřené hlasovým signálem
- **Elektronické rotační** - s přijímačem dálkového ovládání
- **Elektronické rotační ovládané místně tlačítkem**

V některých oblastech jsou umístěny i **místní rozhlas**y, které umožňují vysílání varovného signálu i předání hlasové tísňové informace a jsou zapojeny do systému dálkového ovládání.

Provázanost jednotného systému varování a informování s hromadnými informačními prostředky umožňuje včasné a plošné informování osob o hrozícím nebo vzniklém nebezpečí.

Mluvicí sirény

Aby obyvatelé neprožívali v době zkoušky sirén šok, zda se skutečně něco nestalo, od roku 2009 je u elektronických sirén několik minut před jejich zkouškou odvysílána **verbální**

informace, která na zkoušku sirén upozorní. Tyto informace neumí odvysílat starší typy sirén, tedy rotační. Jsou však postupně nahrazovány moderními. [12]

5.1 Rozbor průběhu elektronických verbálních sirén

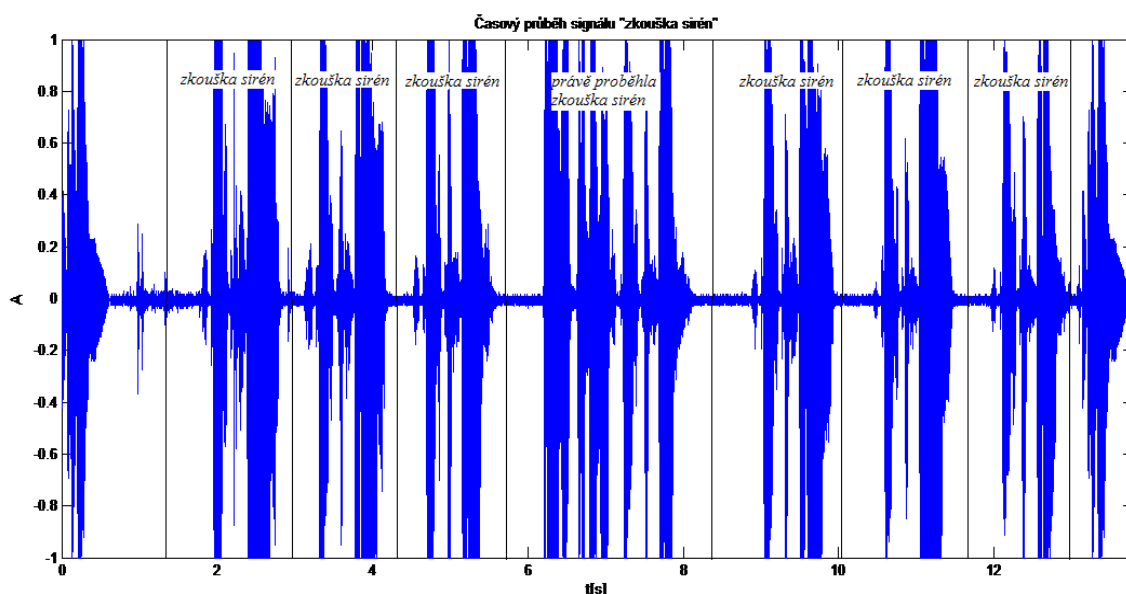
Elektronické verbální sirény patří k moderním typům sirén, u kterých je varovný signál doprovázen také signálem verbálním (hlasovým). V současné době je disponováno se sedmi druhy těchto elektronických verbálních sirén.

Jedná se o [12]:

- Zkouška sirén
- Požární poplach
- Nebezpečí zátopové vlny
- Všeobecná výstraha
- Chemická havárie
- Radiační havárie
- Konec poplachu

Zkouška sirén

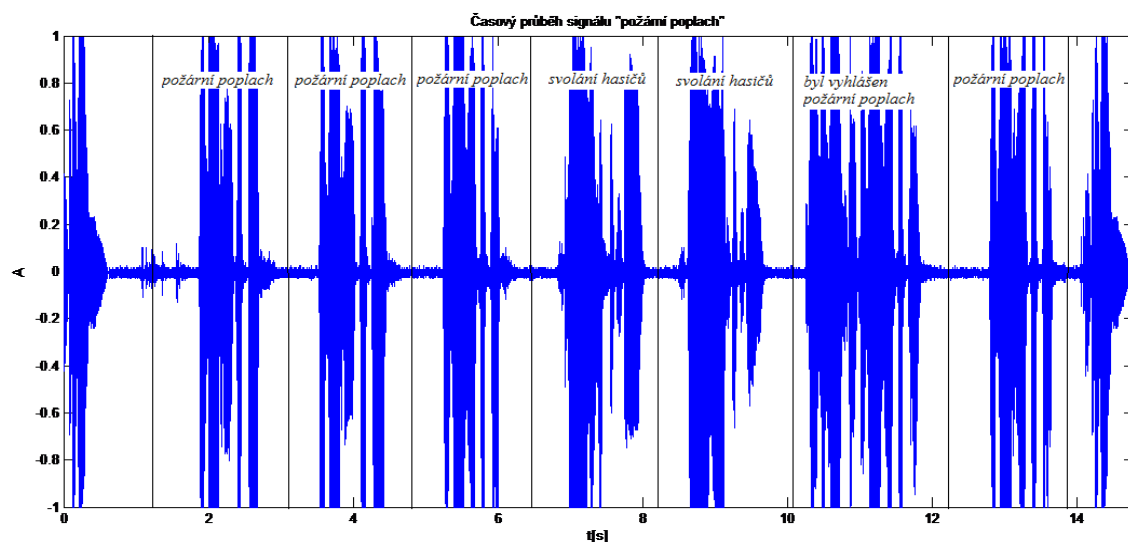
Text: Zkouška sirén, zkouška sirén, zkouška sirén, právě proběhla zkouška sirén, zkouška sirén, zkouška sirén, zkouška sirén.



Obr. 12 Časový průběh signálu verbální sirény - zkouška sirén.

Požární poplach

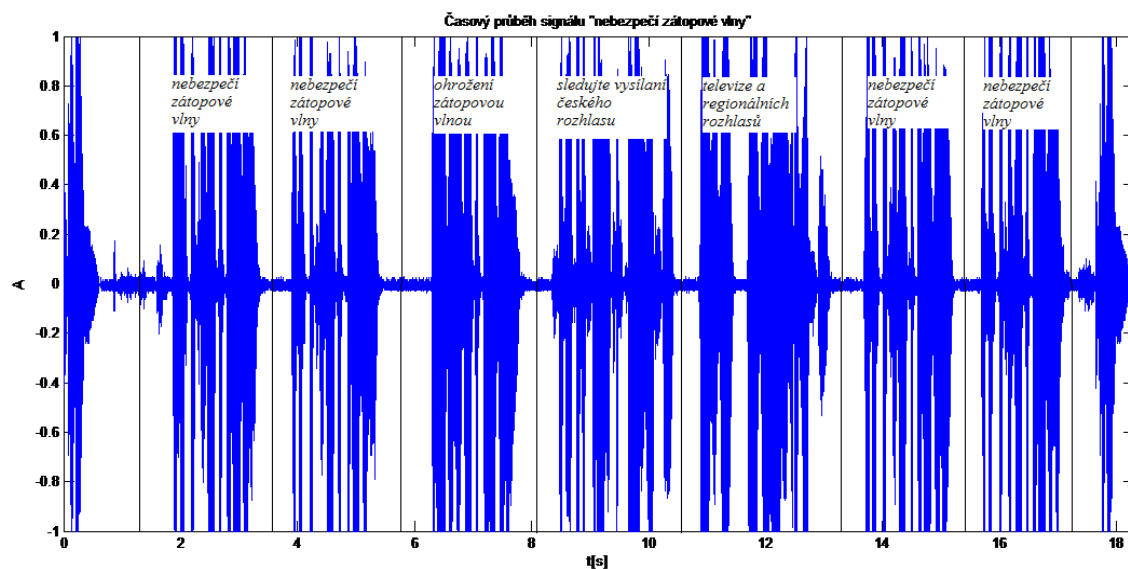
Text: Požární poplach, požární poplach, požární poplach, svolání hasičů, svolání hasičů, byl vyhlášen požární poplach, požární poplach.



Obr. 13 Časový průběh signálu verbální sirény – požární poplach.

Nebezpečí zátopové vlny

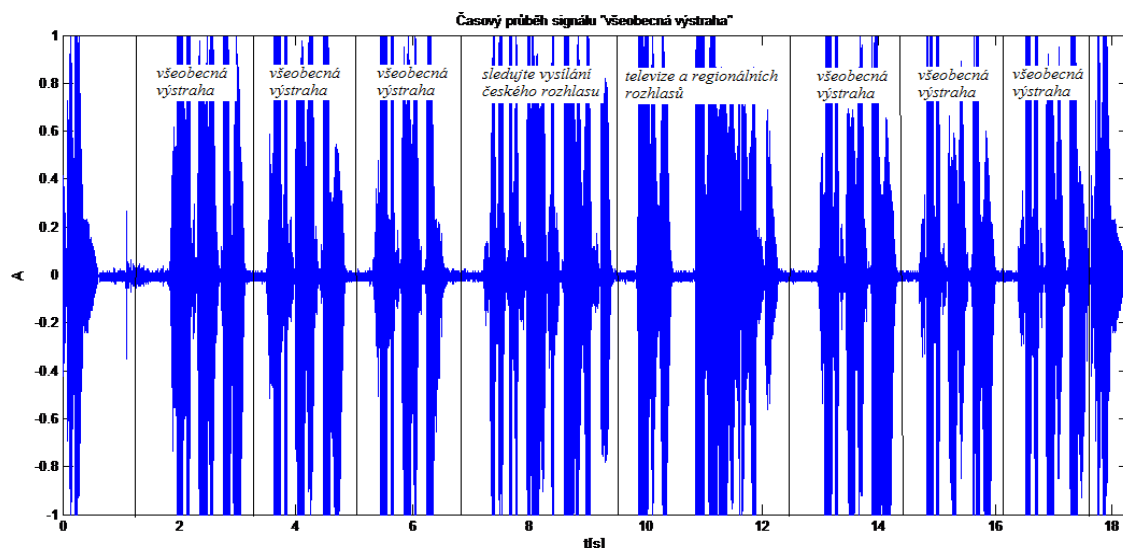
Text: Nebezpečí zátopové vlny, nebezpečí zátopové vlny, ohrožení zátopovou vlnou, sledujte vysílání českého rozhlasu, televize a regionálních rozhlasů, nebezpečí zátopové vlny, nebezpečí zátopové vlny.



Obr. 14 Časový průběh signálu verbální sirény – nebezpečí zátopové vlny.

Všeobecná výstraha

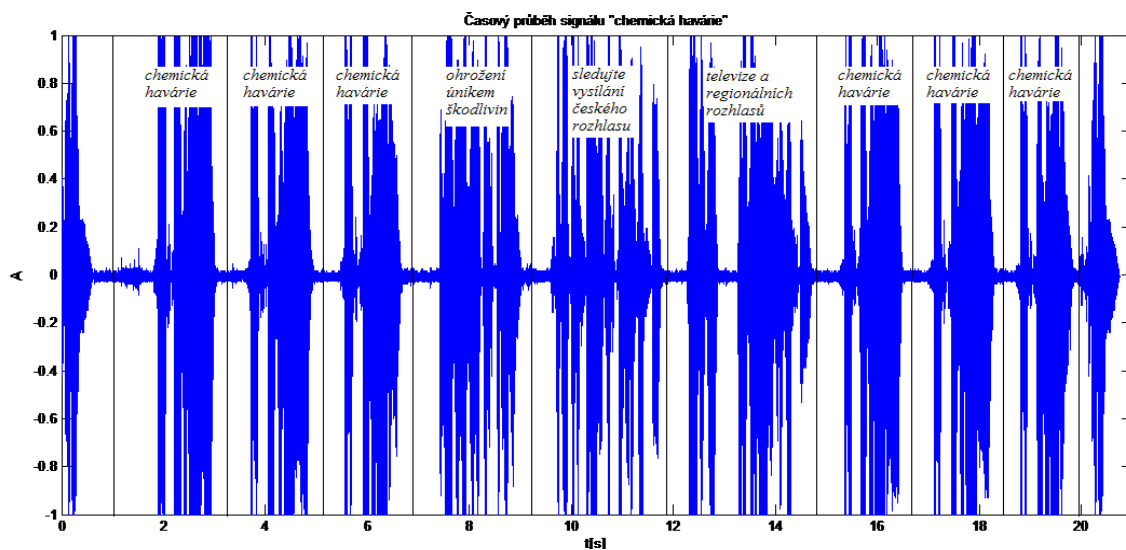
Text: Všeobecná výstraha, všeobecná výstraha, všeobecná výstraha, sledujte vysílání českého rozhlasu, televize a regionálních rozhlasů, všeobecná výstraha, všeobecná výstraha, všeobecná výstraha.



Obr. 15 Časový průběh signálu verbální sirény – všeobecná výstraha.

Chemická havárie

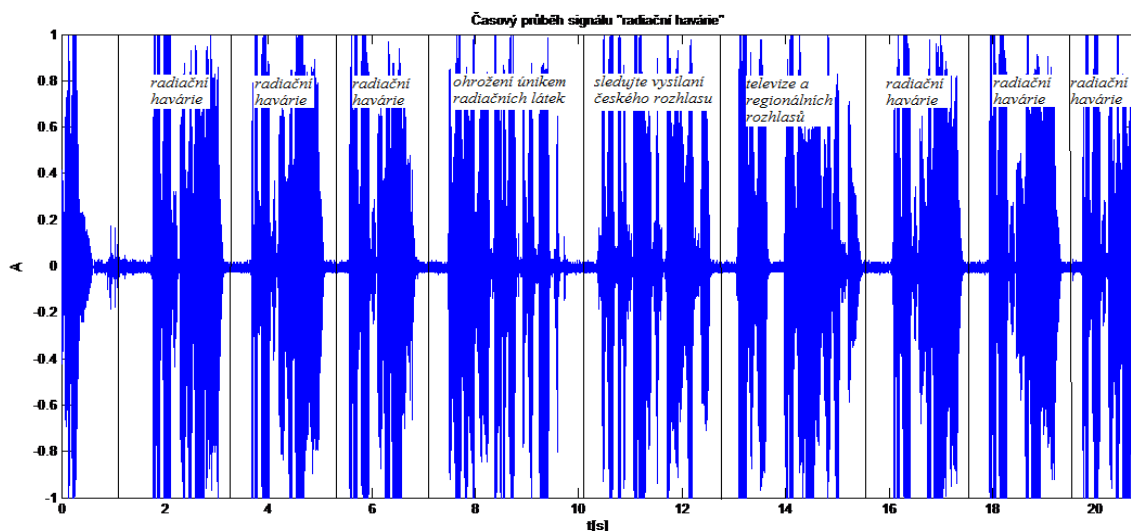
Text: Chemická havárie, chemická havárie, chemická havárie, ohrožení únikem škodlivin, sledujte vysílání českého rozhlasu, televize a regionálních rozhlasů, chemická havárie, chemická havárie, chemická havárie.



Obr. 16 Časový průběh signálu verbální sirény – chemická havárie.

Radiační havárie

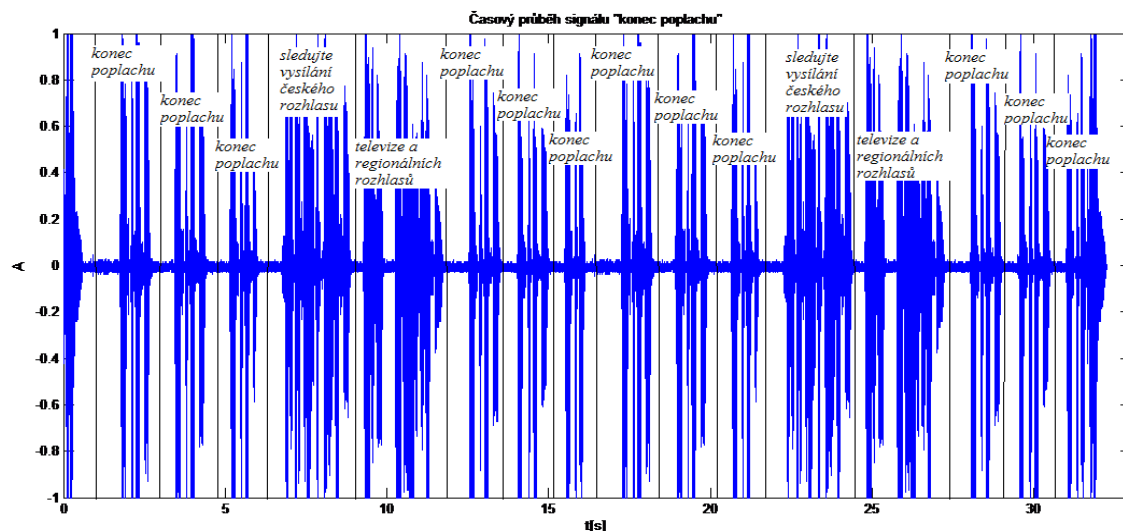
Text: Radiační havárie, radiační havárie, radiační havárie, ohrožení únikem radiačních látek, sledujte vysílání českého rozhlasu, televize a regionálních rozhlasů, radiační havárie, radiační havárie, radiační havárie.



Obr. 17 Časový průběh signálu verbální sirény – radiační havárie.

Konec poplachu

Text: Konec poplachu, konec poplachu, konec poplachu, sledujte vysílání českého rozhlasu, televize a regionálních rozhlasů, konec poplachu, konec poplachu, konec poplachu, konec poplachu, konec poplachu, konec poplachu, sledujte vysílání českého rozhlasu, televize a regionálních rozhlasů, konec poplachu, konec poplachu, konec poplachu.



Obr. 18 Časový průběh signálu verbální sirény – konec poplachu.

Tbl. 1 Délka jednotlivých signálů verbálních sirén.

SIRÉNY	ČASOVÁ DÉLKA ZÁZNAMU [s]	DÉLKA SIGNÁLU VE VZORCÍCH
Zkouška sirén	13,76275	110102
Požární poplach	14,76275	118102
Nebezpečí zátopové vlny	18,26275	146102
Všeobecná výstraha	18,26275	146102
Chemická havárie	20,76275	166102
Radiační havárie	21,76275	174102
Konec poplachu	32,26275	258102

6. Návrh systému pro detekci elektronických verbálních sirén

Před samotným řešením klasifikace varovných verbálních sirén, byla provedena **analýza jednotlivých signálů** převzatých z [12] a to především pro efektivnější stanovení postupu následného praktického řešení.

Do grafů byla vykreslena **střední hodnota rámcovaných signálů**, **střední hodnota signálů po osmi rámcích** a **energie** jednotlivých signálů.

Pro určení středních hodnot byl v **MATLABu** sestaven soubor *mean.m*.

Matlab:

```
Fdata_mean = mean (Fdata) / Lram;
```

A pro výpočet jednotlivých energií soubor *energie.m*.

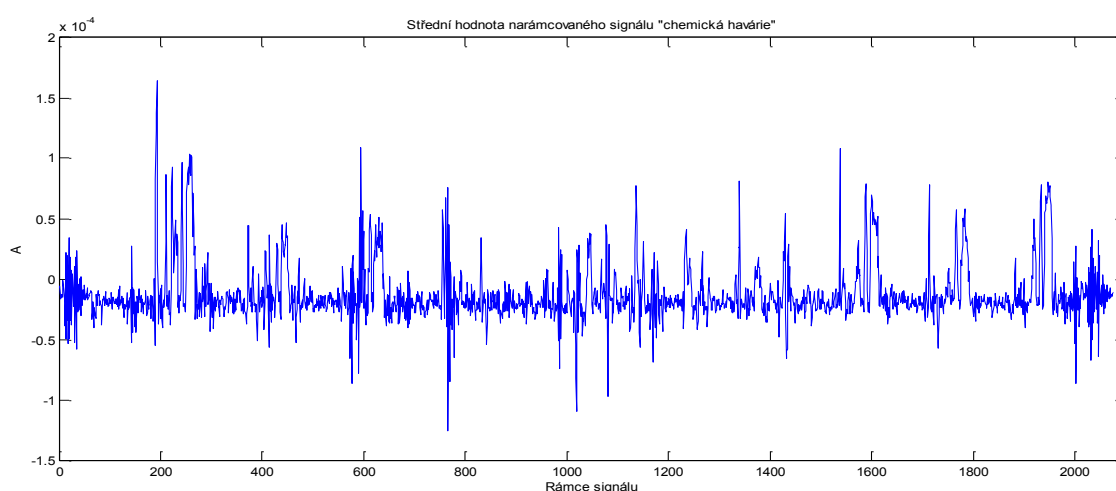
Matlab:

```
E = sum (abs (Fdata).^2) / Lram;
```

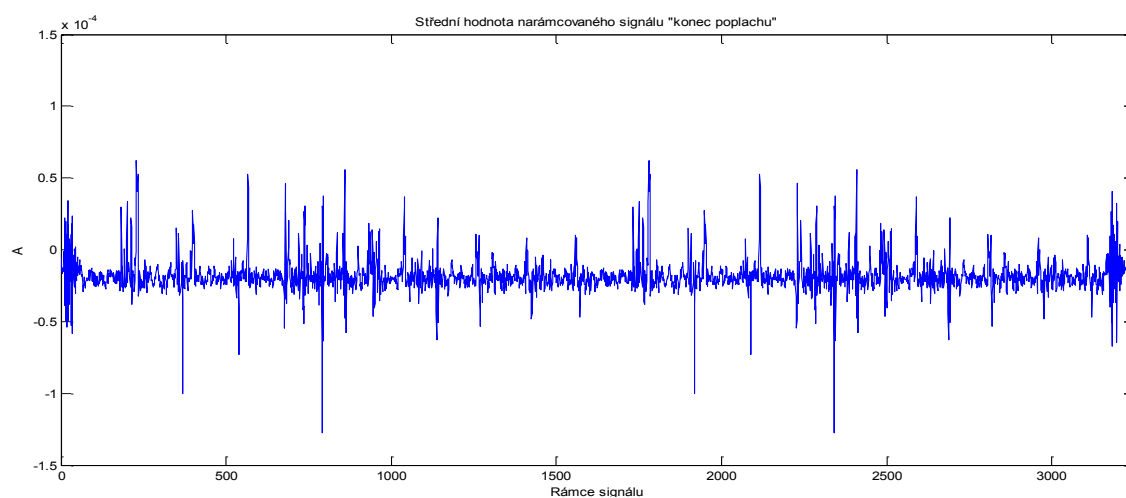
6.1 Zobrazení středních hodnot signálů

V této kapitole můžete vidět grafické výstupy **středních hodnot signálů** elektronických verbálních varovných sirén, ze kterých je patrný jejich odlišný průběh.

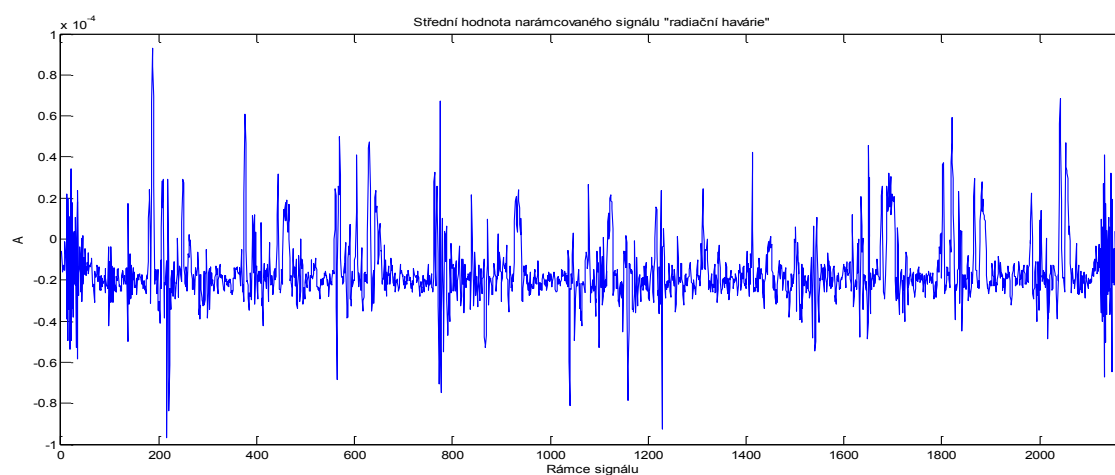
Střední hodnota, jako parametr rozdělení náhodné veličiny, který je definován jako vážený průměr daného rozdělení, je zde vypočtena a vykreslena pro každý rámec jednotlivých signálů.



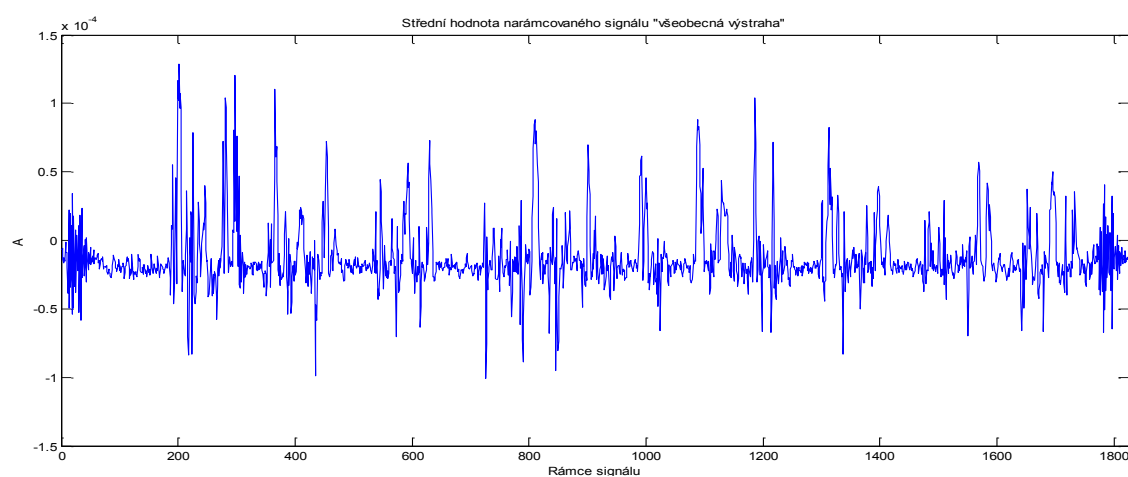
Obr. 19 Střední hodnota rámcovaného signálu – zkouška sirén.



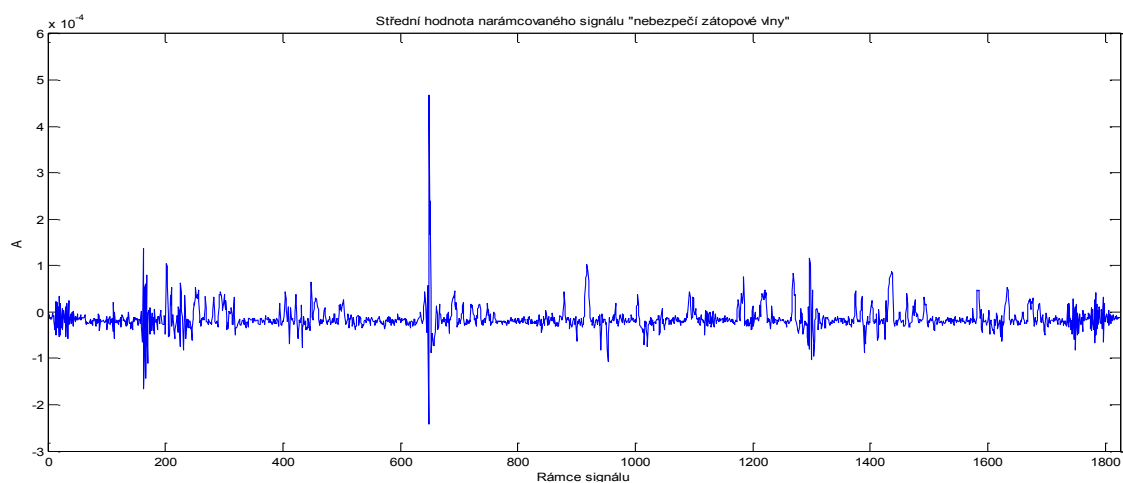
Obr. 20 Střední hodnota rámcovaného signálu – požární poplach.



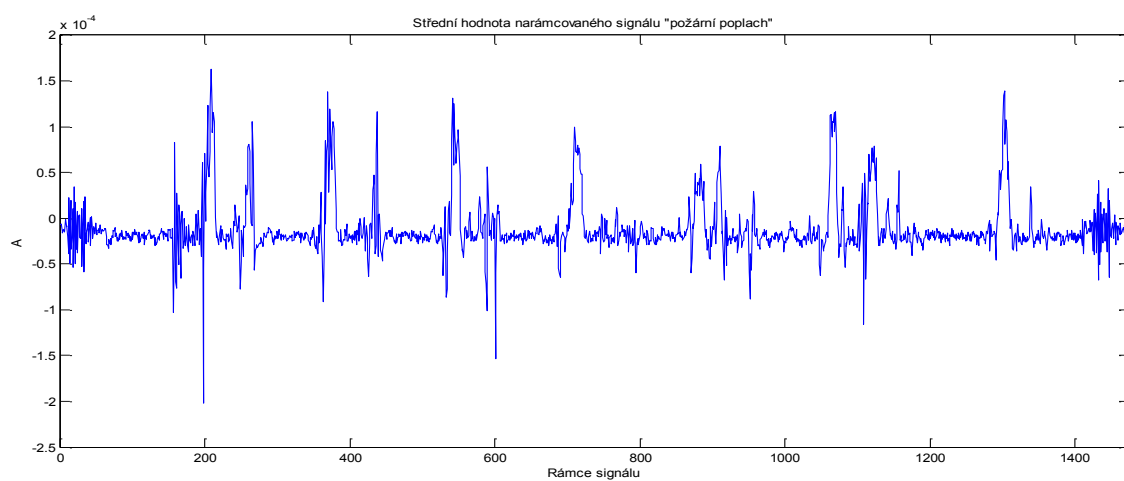
Obr. 21 Střední hodnota rámcovaného signálu – nebezpečí zátopové vlny.



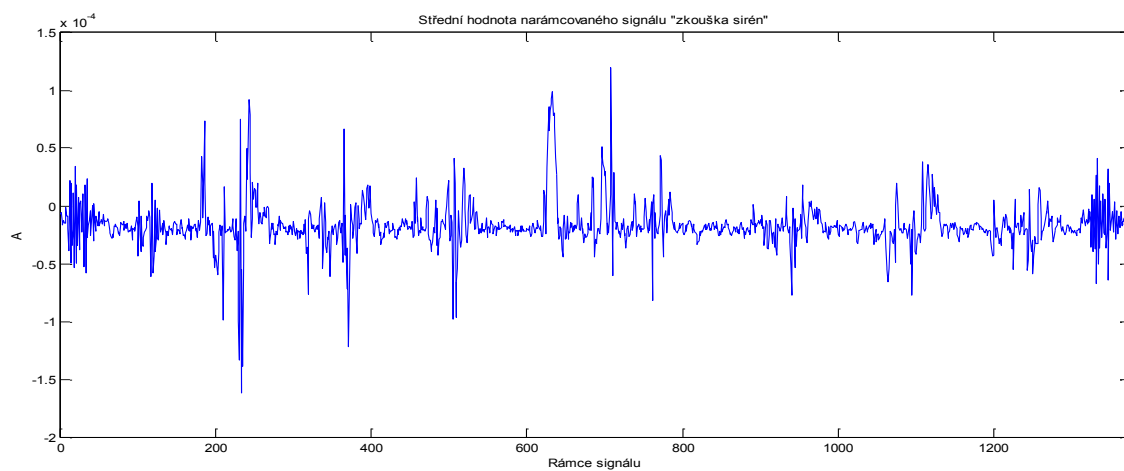
Obr. 22 Střední hodnota rámcovaného signálu – všeobecná výstraha.



Obr. 23 Střední hodnota rámcovaného signálu – chemická havárie.

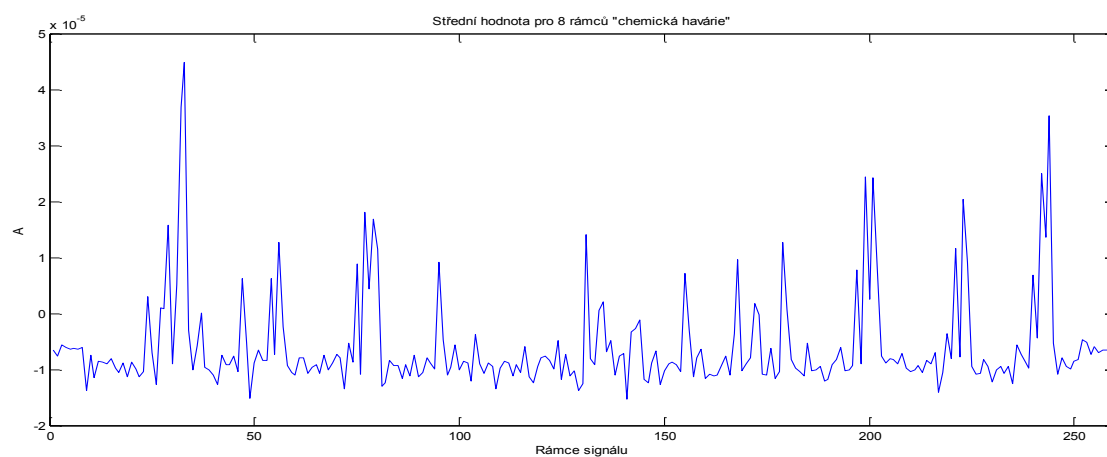


Obr. 24 Střední hodnota rámcovaného signálu – radiační havárie.

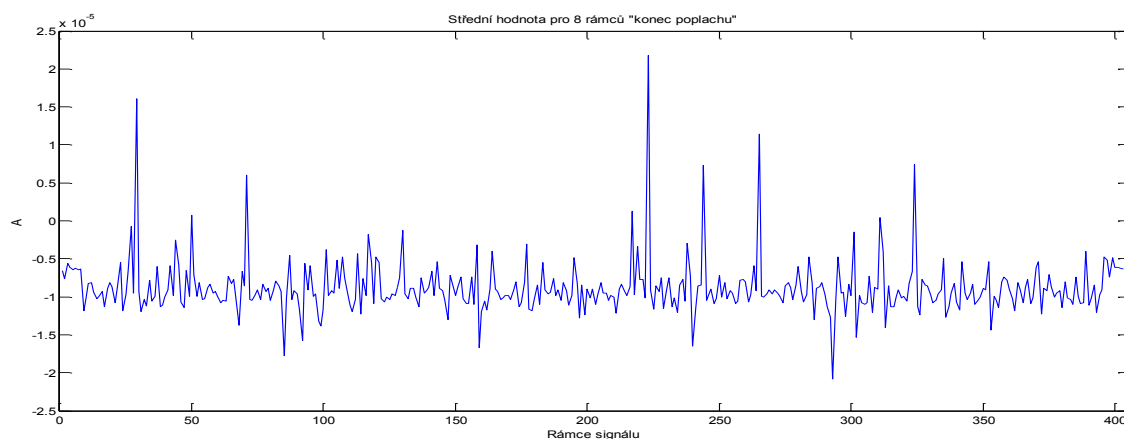


Obr. 25 Střední hodnota rámcovaného signálu – konec poplachu.

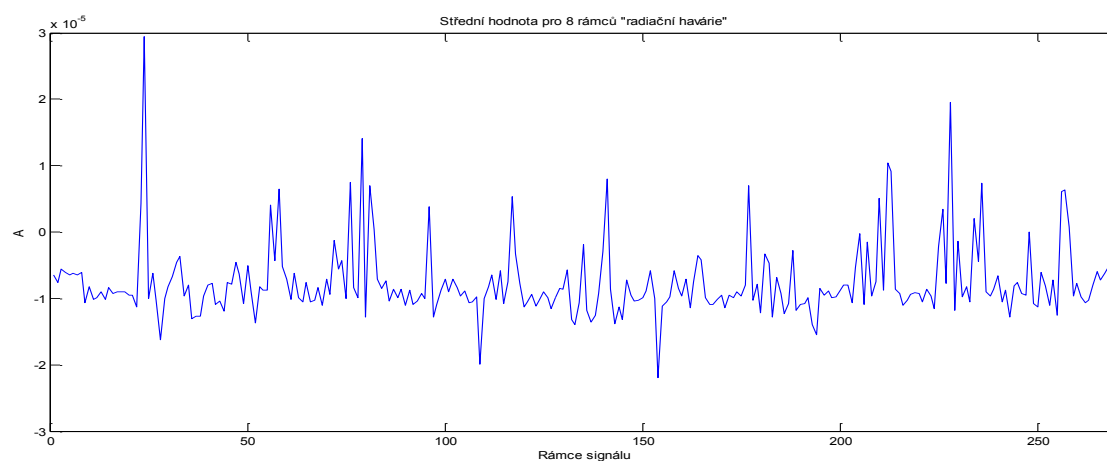
6.1.1 Zobrazení středních hodnot pro 8 rámců



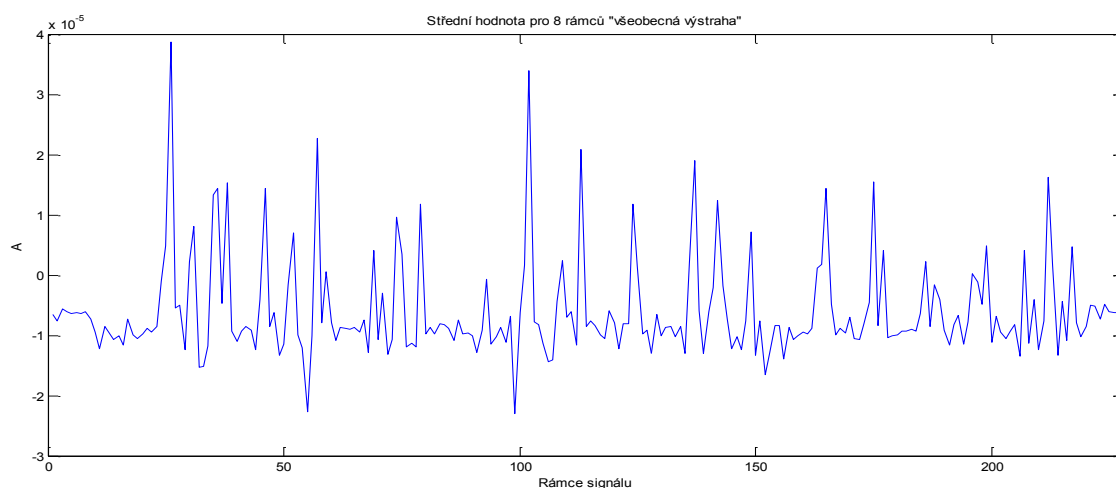
Obr. 26 Střední hodnota signálu po 8-mi rámcích – zkouška sirén.



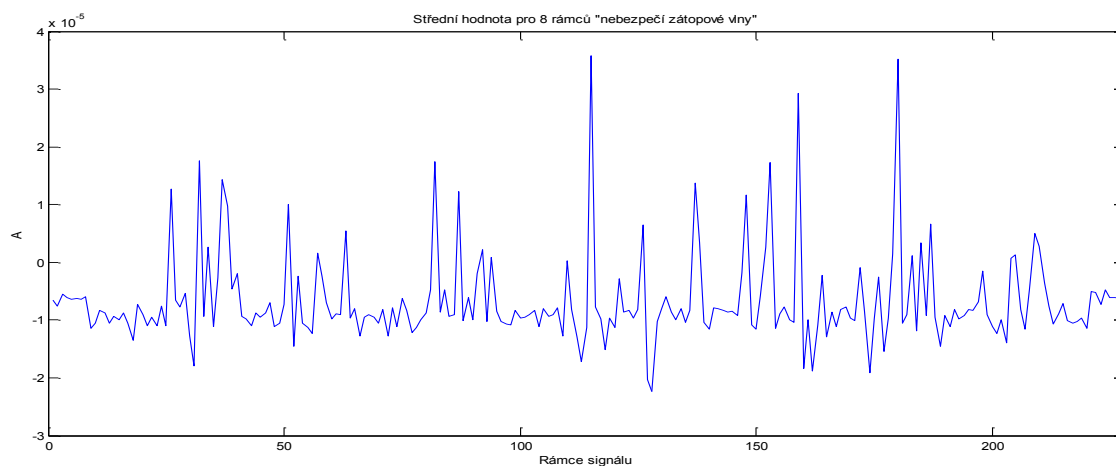
Obr. 27 Střední hodnota signálu po 8-mi rámcích – požární poplach.



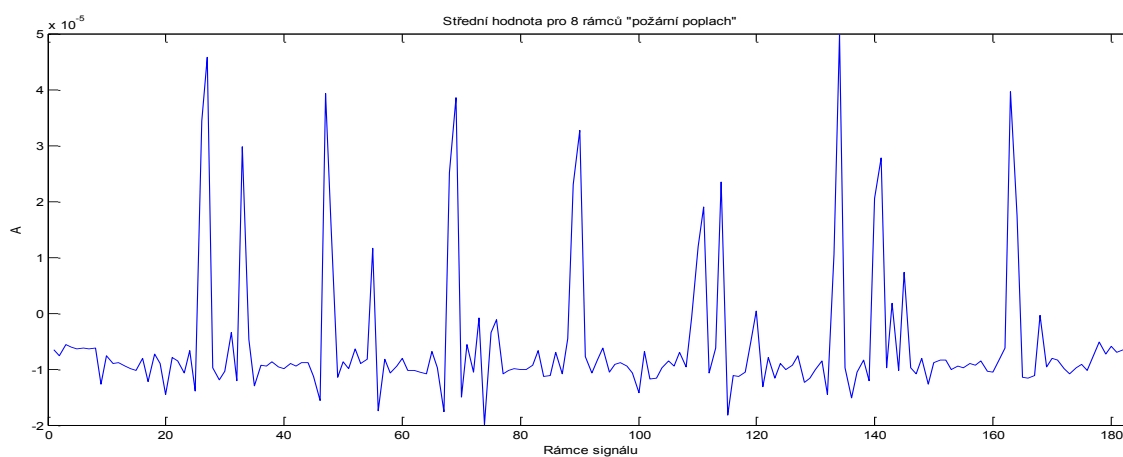
Obr. 28 Střední hodnota signálu po 8-mi rámcích – nebezpečí zátopové vlny.



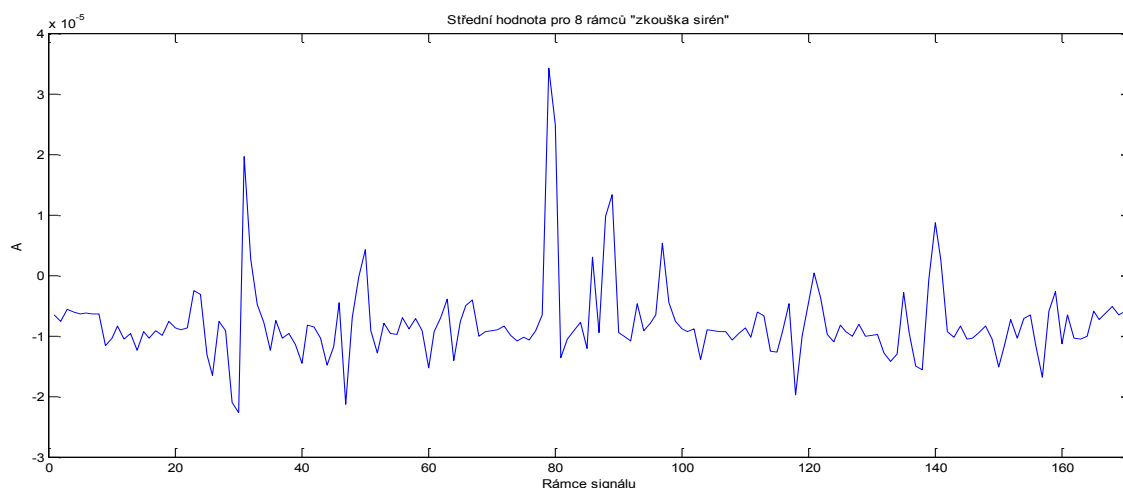
Obr. 29 Střední hodnota signálu po 8-mi rámcích – všeobecná výstraha.



Obr. 30 Střední hodnota signálu po 8-mi rámcích – chemická havárie.



Obr. 31 Střední hodnota signálu po 8-mi rámcích – radiační havárie.

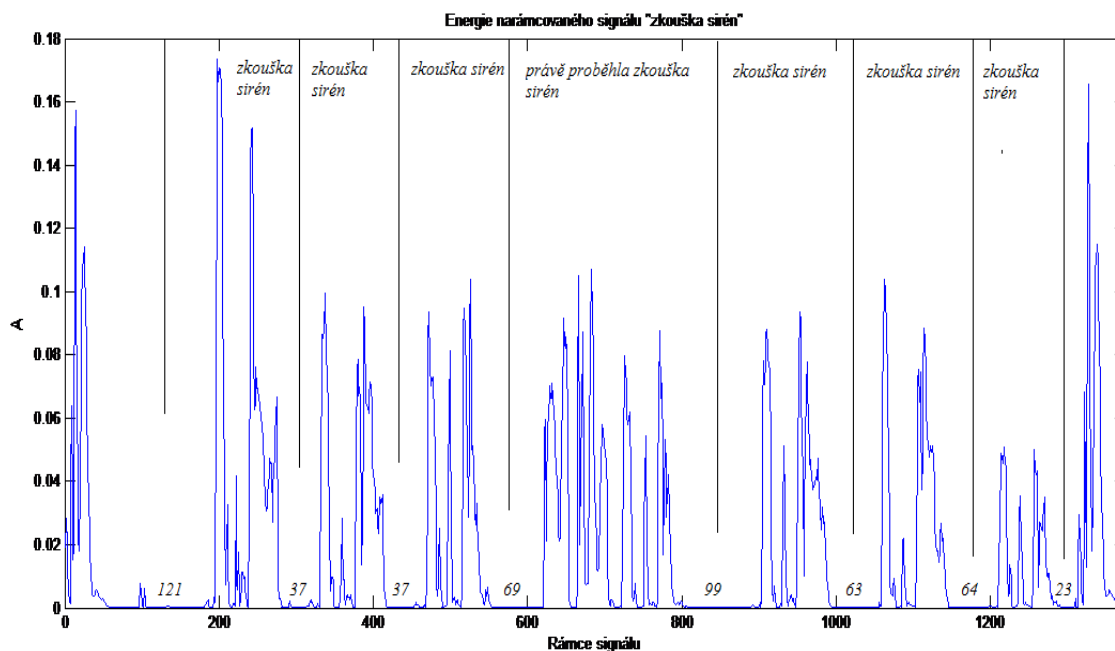


Obr. 32 Střední hodnota signálu po 8-mi rámcích – konec poplachu.

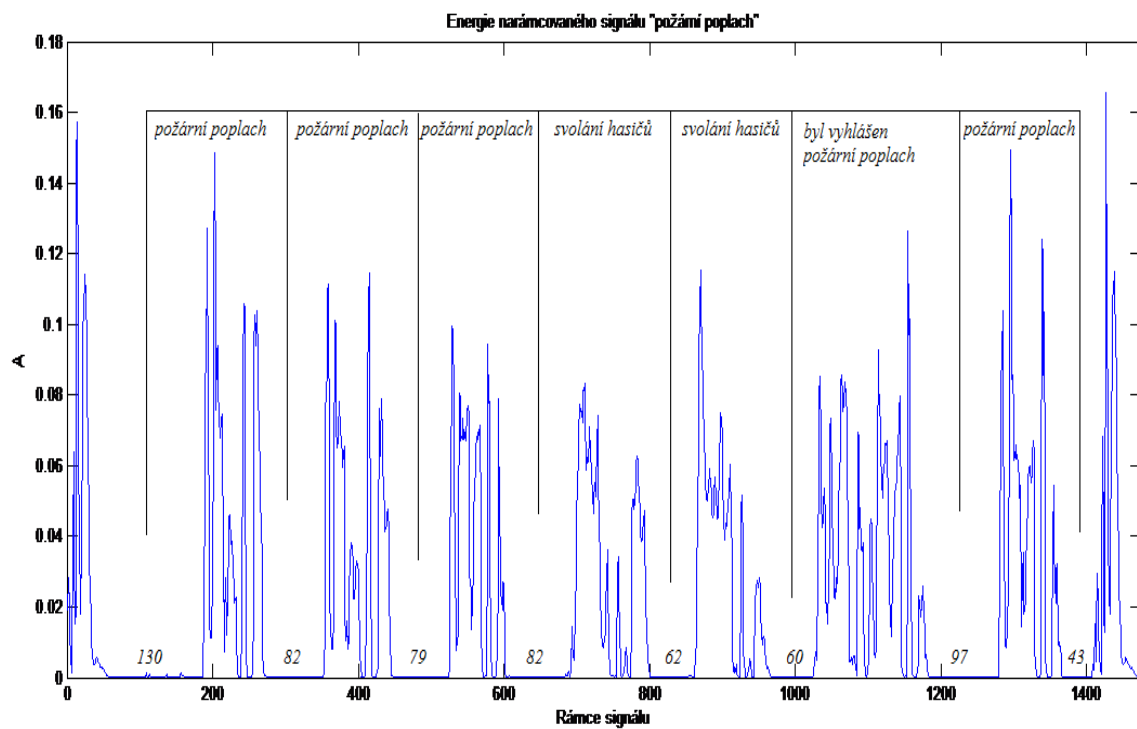
6.2 Zobrazení energie signálů

Na grafických výstupech níže vidíme znázorněnou **energii** pro jednotlivé vzorové signály verbálních sirén.

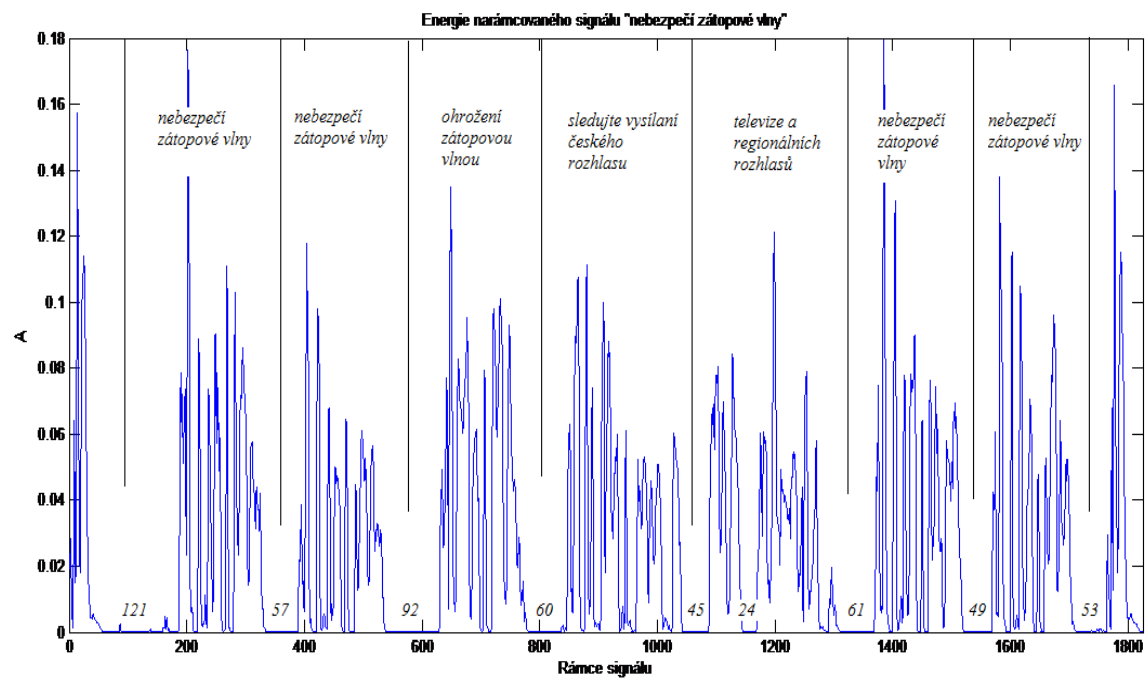
Jsou na nich vyznačeny jednotlivé verbální promluvy i s významem a také počet vzorků v jednotlivých pauzách mezi souslovími, jejichž amplituda dosahuje hodnot 0 až $9,9999 \cdot 10^{-4}$.



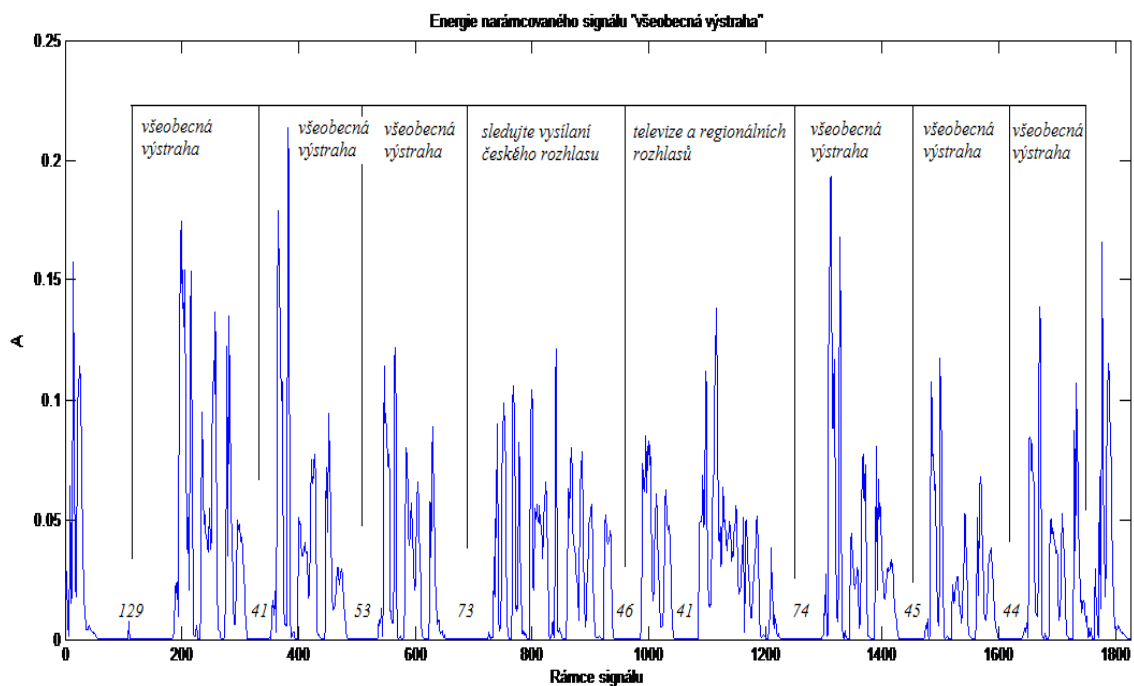
Obr. 33 Energie signálu – zkouška sirén.



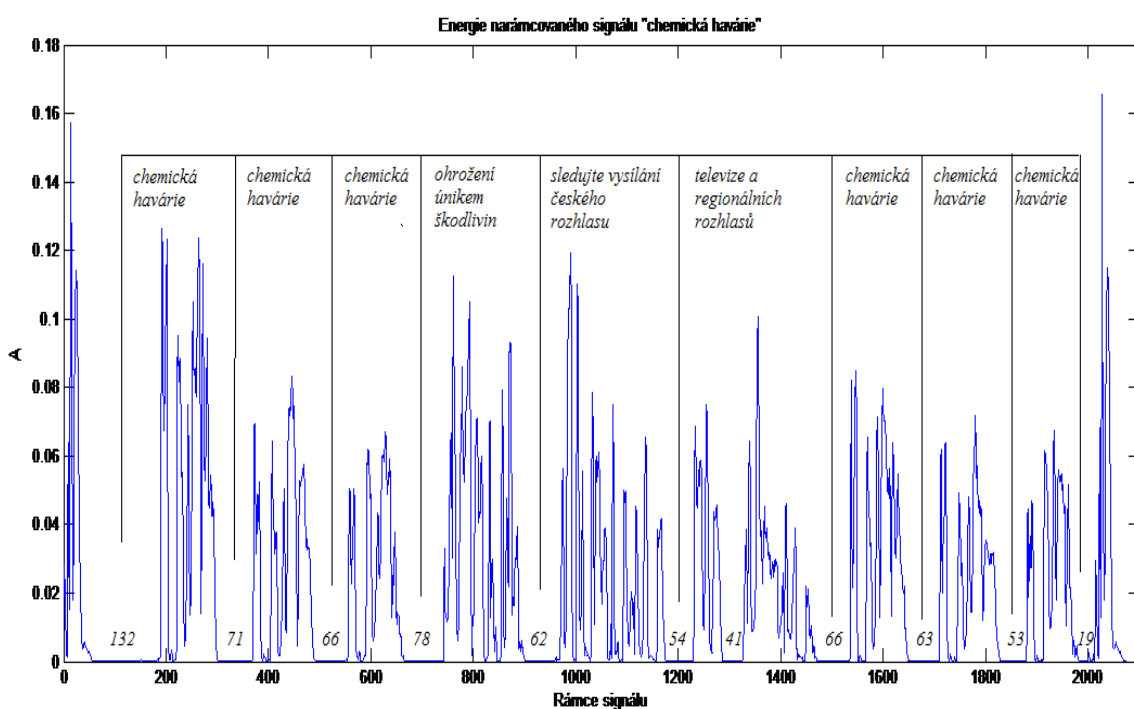
Obr. 34 Energie signálu – požární poplach.



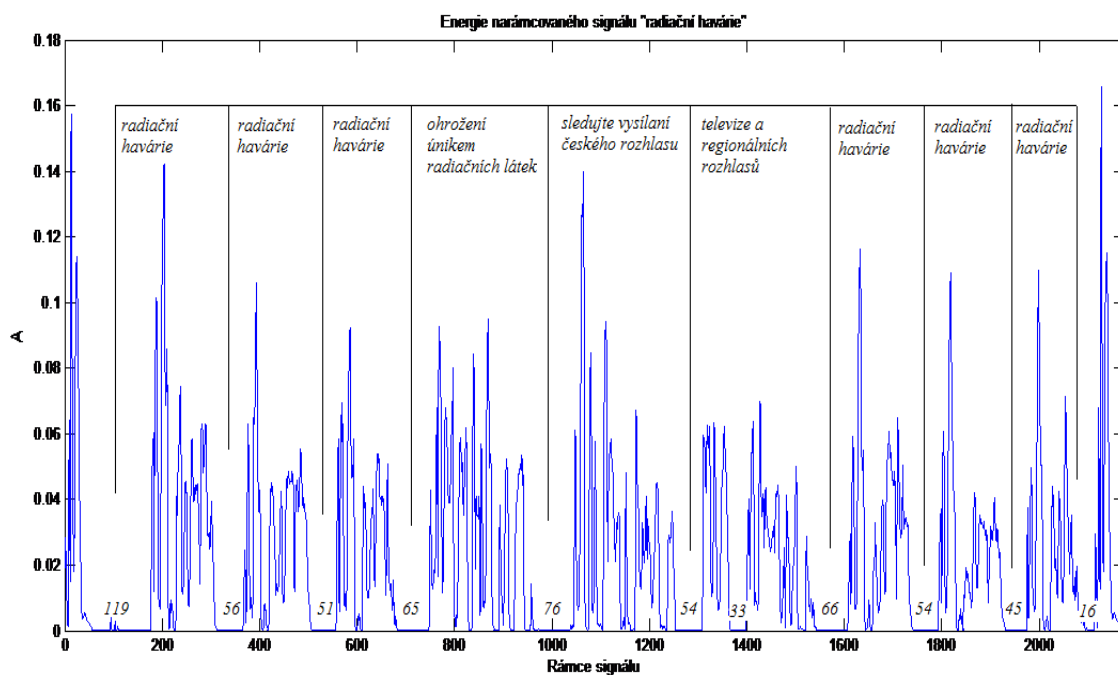
Obr. 35 Energie signálu – nebezpečí zátopové vlny.



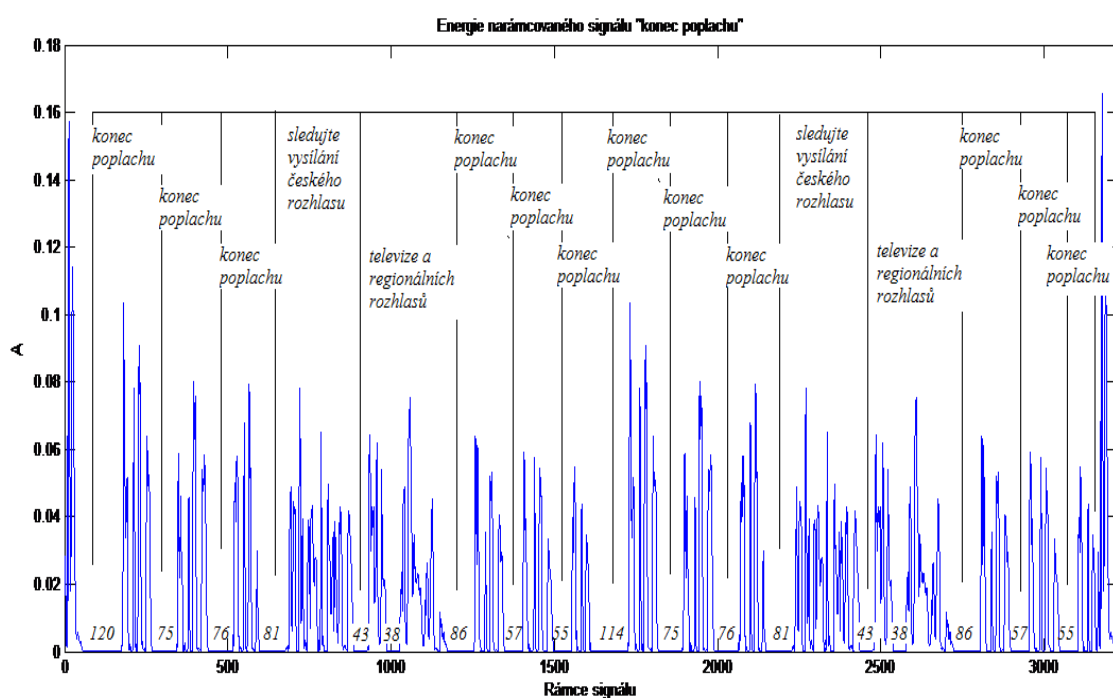
Obr. 36 Energie signálu – všeobecná výstraha.



Obr. 37 Energie signálu – chemická havárie.



Obr. 38 Energie signálu – radiální havárie.



Obr. 39 Energie signálu – konec poplachu.

Po vykreslení energií pro jednotlivé signály byl stanoven **algoritmus pro identifikaci jednotlivých mezer mezi souslovími**. Jedná se o úseky, kde energie dosahuje amplitudy v rozmezí 0 až $9,9999 \cdot 10^{-4}$.

Tento algoritmus byl později také aplikován do hlavního souboru, pro detekci jednotlivých elektronických verbálních sirén, jen s tím rozdílem, že vyhledává hodnoty energie o amplitudě v rozmezí 0 až 10% z maximální hodnoty. Je tomu tak pro přesnější identifikaci v případě razantně zesíleného či zeslabeného signálu.

Pro lokalizaci mezer, výpočet sumy vzorků v jednotlivých mezerách a stanovení časového rozmezí byly použity soubory vytvořené v **MATLABu** s názvem *mezery_1.m – mezery_7.m*.

Výsledky pro jednotlivé vzorové signály jsou přehledně znázorněny v tabulkách Tbl. 2, Tbl. 3, Tbl. 4.

Tbl. 2 Počet vzorků o velikosti v rozmezí $0 - 9,9999 \cdot 10^{-4}$ v jednotlivých pauzách mezi jednotlivými souslovími.

SIRÉNA:	MEZERY:																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Zkouška sirén	121	37	37	69	99	63	64	23										
Požární poplach	130	82	79	82	62	60	97	43										
Nebezpečí zátopové vlny	121	57	92	60	45	24	61	49	42									
Všeobecná výstraha	129	41	53	73	46	41	74	45	44									
Chemická havárie	132	71	66	78	62	54	41	66	63	53	19							
Radiační havárie	119	56	51	65	76	54	33	66	54	45	16							
Konec poplachu	120	75	76	81	43	38	86	57	55	114	75	76	81	43	38	86	57	55

Tbl. 3 Pozice mezer ve vzorcích u jednotlivých vzorových signálů.

MEZERY:	SIRÉNA:						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
1	55 - 181	55 - 186	55 - 186	55 - 185	55 - 186	55 - 176	55 - 174
2	280 - 317	271 - 352	333 - 389	312 - 352	300 - 370	311 - 366	268 - 342
3	417 - 453	446 - 524	538 - 629	484 - 536	489 - 554	506 - 556	439 - 514
4	552 - 620	603 - 684	778 - 837	652 - 724	664 - 741	683 - 747	598 - 678
5	802 - 900	801 - 862	1042 - 1086	940 - 985	900 - 961	963 - 1038	886 - 928
6	993 - 1055	966 - 1025	1145 - 1168	1044 - 1084	1175 - 1228	1254 - 1307	988 - 1025
7	1146 - 1209	1183 - 1279	1308 - 1368	1224 - 1297	1286 - 1326	1365 - 1397	1169 - 1254
8	1287 - 1309	1366 - 1408	1520 - 1568	1428 - 1472	1470 - 1535	1542 - 1607	1347 - 1403
9			1709 - 1763	1598 - 1641	1647 - 1709	1740 - 1793	1498 - 1552
10					1827 - 1879	1930 - 1974	1611 - 1724
11					1981 - 1999	2098 - 2113	1818 - 1892
12							1988 - 2064
13							2148 - 2228

14							2436 - 2478
15							2538 - 2575
16							2719 - 2804
17							2897 - 2953
18							3048 - 3102

Tbl. 4 Pozice mezer v čase t[s] u jednotlivých vzorových signálů.

MEZERY:	SIRENA:						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
1	0,5404 - 1,801	0,5404 - 1,851	0,5403 - 1,851	0,5403 - 1,841	0,5403 - 1,851	0,5402 - 1,751	0,5402 - 1,731
2	2,792 - 3,162	2,702 - 3,512	3,322 - 3,882	3,112 - 3,512	2,991 - 3,692	3,101 - 3,652	2,671 - 3,411
3	4,163 - 4,523	4,453 - 5,234	5,373 - 6,283	4,833 - 5,353	4,882 - 5,533	5,052 - 5,553	4,381 - 5,132
4	5,514 - 6,195	6,024 - 6,835	7,774 - 8,365	6,514 - 7,234	6,633 - 7,404	6,823 - 7,463	5,972 - 6,772
5	8,016 - 8,997	8,005 - 8,616	10,42 - 10,86	9,395 - 9,845	8,994 - 9,658	9,624 - 10,41	8,853 - 9,273
6	9,927 - 10,55	9,657 - 10,25	11,46 - 11,68	10,44 - 10,84	11,75 - 12,28	12,54 - 13,07	9,873 - 10,24
7	11,46 - 12,09	11,83 - 12,79	13,08 - 13,68	12,24 - 12,97	12,86 - 13,26	13,65 - 13,97	11,68 - 12,53

8	12,87 - 13,09	13,66 - 14,08	15,2 - 15,68	14,28 - 14,72	14,67 - 15,35	15,42 - 16,07	13,46 - 14,02
9			17,09 - 17,63	15,98 - 16,41	16,47 - 17,09	17,4 - 17,93	14,97 - 15,51
10					18,27 - 18,79	19,3 - 19,74	16,1 - 17,24
11					19,81 - 20,12	20,98 - 21,13	18,18 - 18,92
12							19,89 - 20,64
13							21,48 - 22,28
14							24,36 - 24,78
15							25,38 - 25,75
16							27,19 - 28,04
17							28,97 - 29,53
18							30,48 - 31,02

7. Realizace systému pro detekci signálu elektronických verbálních sirén

Analýza jednotlivých signálů s praktickým řešením jsou realizovány pomocí programu **MATLAB**. Program byl vytvořen pro konkrétní zadání a nelze ho tedy bez úprav využít pro jiné úkoly.

Celý program je konstruován do jednoho hlavního skriptu. Probíhá v něm celý proces předzpracování i rozpoznávání konkrétních verbálních varovných signálů. V tomto skriptu je také zapracována funkce pro rámcování daného signálu, které je nezbytné pro další výpočty a úpravy.

Některé parametry u procesu předzpracování či zpracování signálu musí být určeny. Hodnoty těchto parametrů byly vybrány většinou na základě doporučení ve studované literatuře nebo z empirických zkušeností.

Tbl. 5 Parametry pro zpracování signálu.

PARAMETRY	HODNOTY
Vzorkovací frekvence	16 kHz
Délka rámce	20 ms
Délka překrytí rámce	10 ms

7.1 Skripty a funkce realizované v programu MATLAB

V této kapitole jsou popsány **funkce a skripty naprogramované pro účely předzpracování a zpracování verbálních signálů varovných sirén.**

7.1.1 Funkce *frame.m*

Tato funkce má za úkol **rozdělit signál na rámce**. Jako vstupní parametry musí být kromě samotného signálu definovány také délka rámce L_{ram} , překryv rámce P_{ram} a vzorkovací frekvence. Délka a překryv jsou zadány v sekundách, funkce si jejich velikosti přepočítá na počet vzorků. Pomocí funkce *frame* vzniká z vektoru vstupního signálu matice, kde sloupce značí jednotlivé rámce a řádky počet vzorků v rámci.

Rámcování se nejčastěji dělá za pomoci Hammingova okna. Ve funkci to znamená, že vzorky v rámci vynásobíme Hammingovým oknem, pro které je v MATLABu předdefinovaná funkce.

Výstupními parametry funkce, krom již zmíněné matice vzorků, je délka rámce ve vzorcích, která je důležitá pro další výpočetní procesy.

Matlab:

```
function [matice_vzorku, Lram] = frame (data, delka_ramce, Pram, Fvz)

delka_signalu = length (data);

Lram = delka_ramce * Fvz;

Pram = Pram * Fvz;

posuv_ramce = Lram - Pram;

sloupce_matice = floor (1+((delka_signalu - Lram)/posuv_ramce)); % stanovíme počet sloupců matice

matice_vzorku = zeros (Lram, sloupce_matice); % inicializace matice

% cyklus, kde každému sloupci přiřadíme jeden rámec

for k = 1 : 1 : sloupce_matice % pro každý sloupec

    for l = 1 : 1 : Lram % pro každý řádek

        if ((k - 1) * posuv_ramce + l) <= delka_signalu % dokud je z čeho načítat a nedosáhneme konce signálu

            matice_vzorku (l, k) = data ((k - 1) * posuv_ramce + l); % plníme matici vzorky ze signálu

        else matice_vzorku (l, k) = 0; % zbytek do konce sloupce 0

        end;

    end;

end;

for k = 1 : 1 : sloupce_matice % pro všechny rámce

    matice_vzorku (1 : Lram, k) = matice_vzorku (1 : Lram, k).*(hamming(Lram)); % rámce vynásobíme Hammingovým oknem

end
```

7.1.2 Skript klasifikace_signalu.m

V tomto souboru je obsažena jak část předzpracování, tak samotné **rozpoznání modifikovaného verbálního signálu** (rušení, posun, zesílení, zeslabení) na základě podobnosti se signálem vzorovým, jehož průběh byl převzat z [12] a analyzován v předchozích kapitolách.

Na začátku skriptu je připravená načítací funkce pro soubory ve formátu:

- wav

Matlab:

```
[data_vzor,Fvz,Nbits]= wavread('pozarni_poplach.wav');
```

Poté je zakomponována funkce rámcování **frame.m**, která signál upraví do matice, pro další zpracování.

Matlab:

```
[Fdata,Lram] = frame(data, 0.02, 0.01, Fvz);
```

„V rámci práce byly vytvořeny i další skripty, díky kterým můžeme zobrazit pro jednotlivé rámce například frekvenční charakteristiku či výkonovou spektrální hustotu (viz teoretická část práce).“

Pro úspěšné vyhodnocení je však důležitý výpočet energie jednotlivých rámců jak vzorového, tak modifikovaného verbálního signálu.

Matlab:

```
E = sum (abs (Fdata).^2) / Lram;
```

Na základě tohoto výpočtu a dalších algoritmů dochází v rámci programu k lokalizaci mezer a stanovení sumy vzorků v daných mezerách mezi jednotlivými souslovími všech signálů (klasifikovaného i vzorových), kde hodnoty energie dosahují téměř nulových hodnot, či hodnot v intervalu od 0 do 10% z maximální hodnoty energie signálu.

Dle předem daného vzorce dochází k výpočtu procentuální podobnosti daných signálu (na základě počtu vzorků v mezerách) a k vyhodnocení výsledků. Každá detekce je opatřena podmínkami, při jejichž nesplnění nedojde ke klasifikaci daného signálu.

V závěru celého souboru, jako součást vyhodnocení je zpracováno GUI sloužící pro přehlednější výčet výsledků.

7.2 Uživatelské rozhraní softwarové aplikace

Pro názornou ukázkou a zřetelnější popis celého procesu vyhodnocení byl zvolen signál verbální sirény „pozarni_poplach.wav“, upravený pomocí algoritmu:

Matlab:

% posun signálu o +3200 vzorků podél časové osy

```
data_vzor_posun = [zeros(1, 3200) data_vzor'];
```

```
data_posun = data_vzor_posun(1 : N);
```

```
data_posun = data_posun';
```

% rušení signálu (data + náhodná data se směrodatnou odchylkou 0.15)

```
ruseni = random('normal', 0, 0.15, N, 1);
```

% zesílení o 1,5 krát

```
data = 1.5 * (data_posun + ruseni);
```

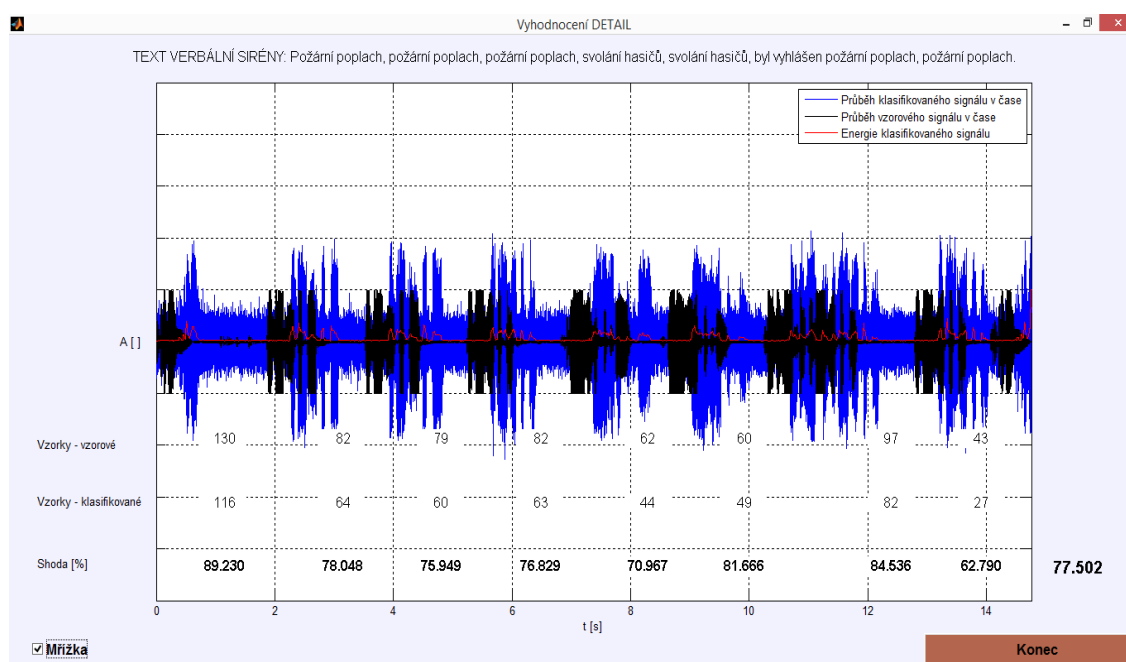
Výsledkem byla ± 78 % shoda se vzorovým signálem „pozarni_poplach.wav“.



Obr. 40 Klasifikační okno, zobrazí se při rozpoznání signálu verbální varovné sirény.

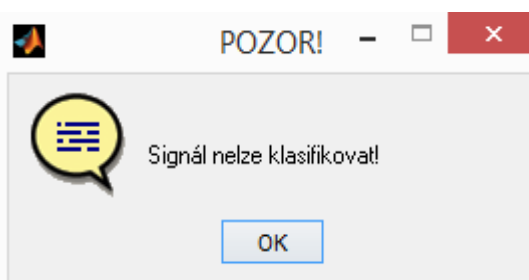
Na shodu nás upozorní okno „klasifikace signálu“, kde se dočteme, o jaký verbální signál se jedná a jaká je procentuální shoda se signálem vzorovým. Dále je zde také uveden text verbální sirény a parametry klasifikovaného signálu (čas, počet vzorků a vzorkovací frekvence).

Máme možnost podívat se také na „detail vyhodnocení“, který obsahuje časový průběh klasifikovaného signálu (modifikovaný signál dle algoritmu), časový průběh vzorového signálu a průběh energie klasifikovaného signálu, která je důležitou součástí pro samotné vyhodnocení. Dále zde také vidíme počet vzorků v mezerách vzorového signálu, počet vzorků v mezerách klasifikovaného signálu, procentuální shodu jednotlivých mezer a celkovou procentuální shodu těchto dvou signálů.



Obr. 41 Detail vyhodnocení.

V případě abnormálních změn hodnoceného signálu (velký šum, posun), kdy nejsou splněny podmínky (více výsledků bylo shodných či vyšší procento shody bylo zaznamenáno u nesprávného signálu) a signál tudíž nelze rozeznat, nás program upozorní na danou situaci výstražným oknem „pozor“.



Obr. 42 Okno POZOR! Zobrazí se v případě, že signál nelze klasifikovat.

Na takto založeném principu rozeznávání bylo provedeno několik měření dle náhodně zvolených algoritmů.

8. Vizualizace a srovnání naměřených výsledků

V této kapitole jsou **zaznamenány a vyhodnoceny veškeré výsledky**, ke kterým jsem při aplikaci vytvořeného softwarového programu dospěla.

Jednotlivé výstupy byly zapsány do přehledných tabulek. Pro lepší orientaci je znázorněna tabulka Tbl. 6, která popisuje jednotlivé části použitých výsledných tabulek.

Tbl. 6 Tabulka popisující jednotlivé části výsledných tabulek.

VZOROVÉ SIGNÁLY:	Algoritmus pro úpravu (modifikaci) signálu						
	Upravený signál	Upravený signál	Upravený signál	Upravený signál	Upravený signál	Upravený signál	Upravený signál
Vzorový signál	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]
Vzorový signál	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]
Vzorový signál	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]
Vzorový signál	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]
Vzorový signál	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]
Vzorový signál	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]
Vzorový signál	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]	Shoda [%]

Každá „**výsledná tabulka**“ obsahuje výsledky měření pro všechny signály verbálních sirén upravených dle daného algoritmu a jejich procentuální shodu se všemi vzorovými signály verbálních sirén, které byly převzaty z [12].

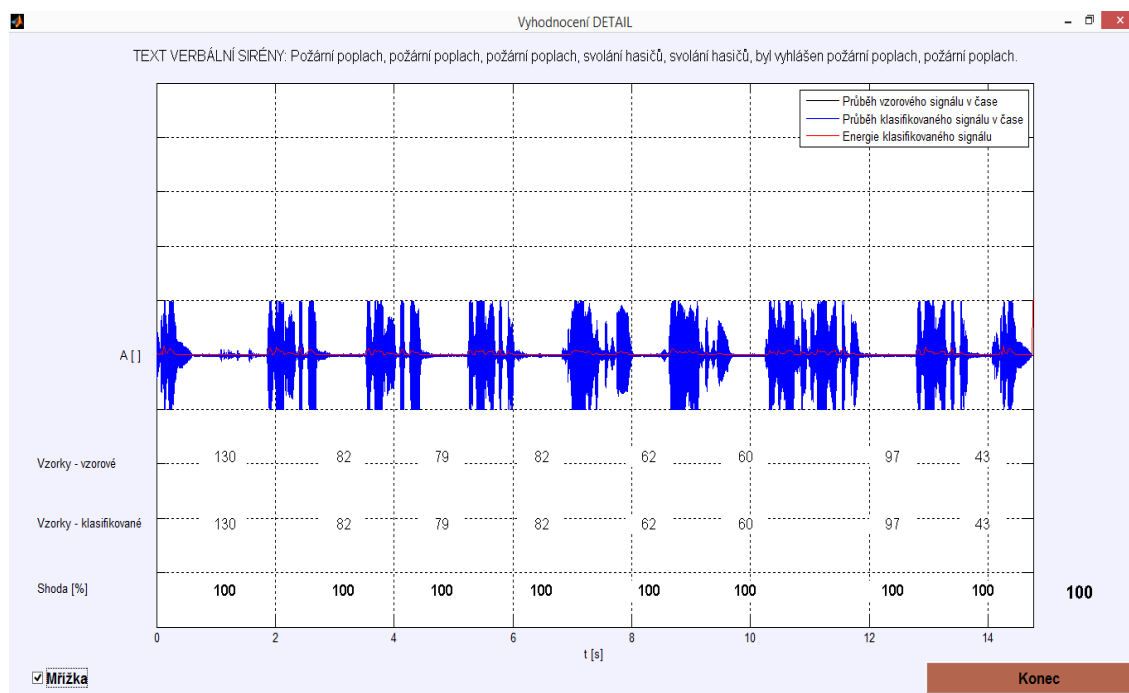
Procentuální shoda je uvedena v [%]. Tam, kde tato shoda dosahuje nejvyšších hodnot, došlo k rozpoznání klasifikovaného signálu. V případě, že nebyly splněny podmínky měření (více výsledků bylo shodných či vyšší procento shody bylo zaznamenáno u nesprávného signálu), je v políčku zapsáno „nelze klasifikovat“. Tyto podmínky slouží především k zabránění nesprávné klasifikace verbální sirény.

Pod každou tabulkou je také uveden jako příklad „**detail vyhodnocení**“ modifikovaného signálu „**pozarni_poplach.wav**“, pro lepší představu a orientaci v dosažených výsledcích.

Pro začátek bylo uskutečněno měření zcela neupraveného signálu, tudíž měření signálu totožného se vzorovým Tbl. 7 (průběhy v grafu jsou shodné – klasifikovaný signál překrývá signál vzorový). Z výsledků tohoto měření lze vyčíst vzájemná procentuální shoda jednotlivých vzorových signálů.

Tbl. 7 Výsledky klasifikace totožného signálu v [%].

VZOROVÉ SIGNÁLY:	SIGNÁLY BEZ ÚPRAV (totožné se vzorovými):						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	100	69.8175	70.3552	59.8272	66.3904	67.3648	81.6271
Požární poplach	70.8588	100	56.6226	71.0382	60.6203	67.6591	88.4201
Nebezpečí zátopové vlny	65.1110	48.9799	100	64.5866	64.5017	69.8961	74.4994
Všeobecná výstraha	53.9198	69.6929	60.6749	100	76.0752	74.4507	72.7467
Chemická havárie	49.3040	44.7538	44.8078	63.9838	100	68.3877	86.5407
Radiační havárie	45.2134	49.4045	54.8188	63.6144	58.2468	100	75.4422
Konec poplachu	34.8059	39.7298	30.3487	38.1020	45.7335	41.8338	100



Obr. 43 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

8.1 Vliv zesílení a zeslabení klasifikovaného signálu na detekci

Při **zesílení** vzorového signálu 20 krát, můžeme z měření níže pozorovat 100% shodu u všech signálů. Stejně, jak tomu bylo u signálů totožných.

Snadno lze odvodit, že samotné zesílení klasifikovaných signálů nemá na správnost výsledku nijak extrémní vliv.

Signály sirén byly mimo dané měření zesíleny také 1000 krát, a správnost výsledků nebyla nijak ovlivněna.

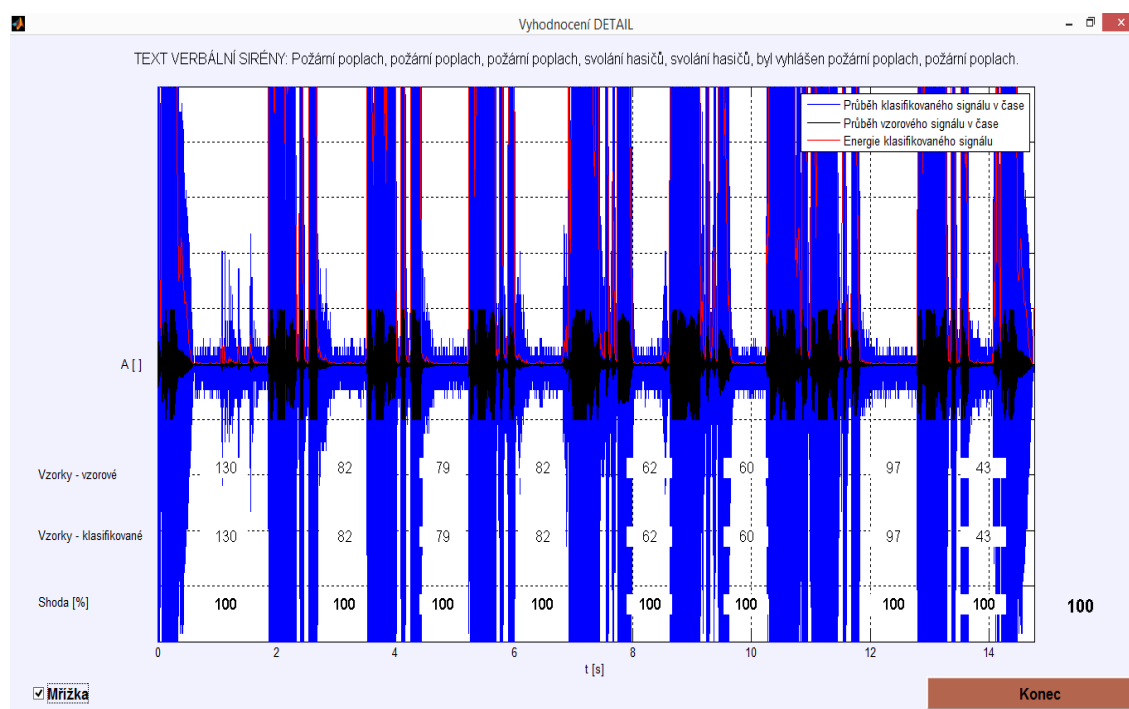
Tbl. 8 Výsledky klasifikace modifikovaného signálu (zesílení 20 krát) v [%].

VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: zesílení *20						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	100	69.8175	70.3552	59.8272	66.3904	67.3648	81.6271

Požární poplach	79.8704	100	56.6226	71.0382	60.6203	67.6591	88.4201
Nebezpečí zátopové vlny	87.3332	71.2021	100	64.5866	64.5017	69.8961	74.4994
Všeobecná výstraha	76.1420	80.8040	60.6749	100	76.0752	74.4507	72.7467
Chemická havárie	85.6676	80.0155	62.8181	81.9941	100	68.3877	86.5407
Radiační havárie	87.0867	85.7681	73.0006	81.7962	67.3377	100	75.4422
Konec poplachu	93.0905	95.2854	74.0524	81.8057	79.0669	73.1781	100

Matlab:

$data = data_vzor * 20;$



Obr. 44 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

Při extrémním **zeslabení** signálů 0,01 krát, byla také zaznamenána 100% shoda u všech klasifikovaných signálů. Znamená to, že zeslabení signálů na důvěryhodnost výsledků, nemá žádný fatální vliv.

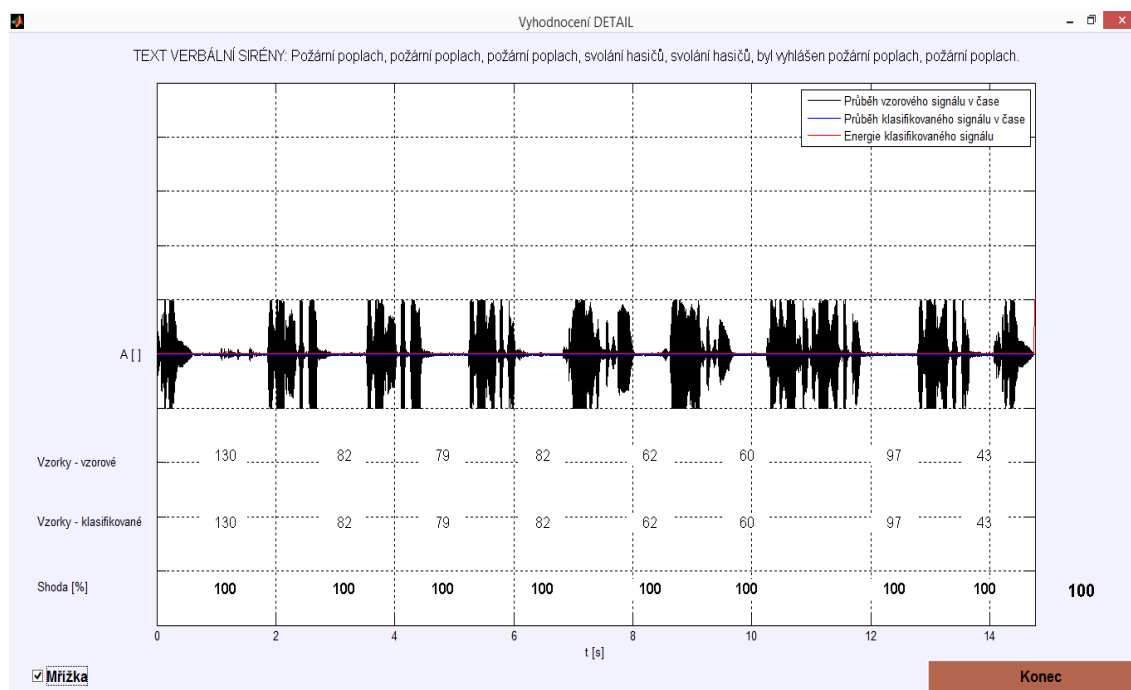
Signály sirén byly mimo dané měření zeslabeny také 0,001 krát, a správnost výsledků nebyla nijak ovlivněna.

Tbl. 9 Výsledky klasifikace modifikovaného signálu (zeslabení 0,01 krát) v [%].

VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: zeslabení *0.01						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	100	69.8175	70.3552	59.8272	66.3904	67.3648	81.6271
Požární poplach	70.8588	100	56.6226	71.0382	60.6203	67.6591	88.4201
Nebezpečí zátopové vlny	65.1110	48.9799	100	64.5866	64.5017	69.8961	74.4994
Všeobecná výstraha	53.9198	69.6929	60.6749	100	76.0752	74.4507	72.7467
Chemická havárie	49.3040	44.7538	44.8078	63.9838	100	68.3877	86.5407
Radiační havárie	45.2134	49.4045	54.8188	63.6144	58.2468	100	75.4422
Konec poplachu	34.8059	39.7298	30.3487	38.1020	45.7335	41.8338	100

Matlab:

*data = data_vzor * 0.01;*



Obr. 45 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

8.2 Vliv rušení klasifikovaného signálu na detekci

Z předchozího měření jsme se dozvěděli, že samotné zeslabení či zesílení nemají žádný vliv na 100% shodnou klasifikaci jednotlivých signálů.

Naopak **silné rušení** už výsledky měření viditelně ovlivňuje. Při přičtení náhodných dat se směrodatnou odchylkou 0,25 se podařilo klasifikovat všechny signály, ale procentuální shoda už není nijak vysoká, dosahuje v průměru 18%.

Při vyšších směrodatných odchylkách se podařilo klasifikovat jen některé signály verbálních sirén nebo už žádné.

Tbl. 10 Výsledky klasifikace modifikovaného signálu (rušení) v [%].

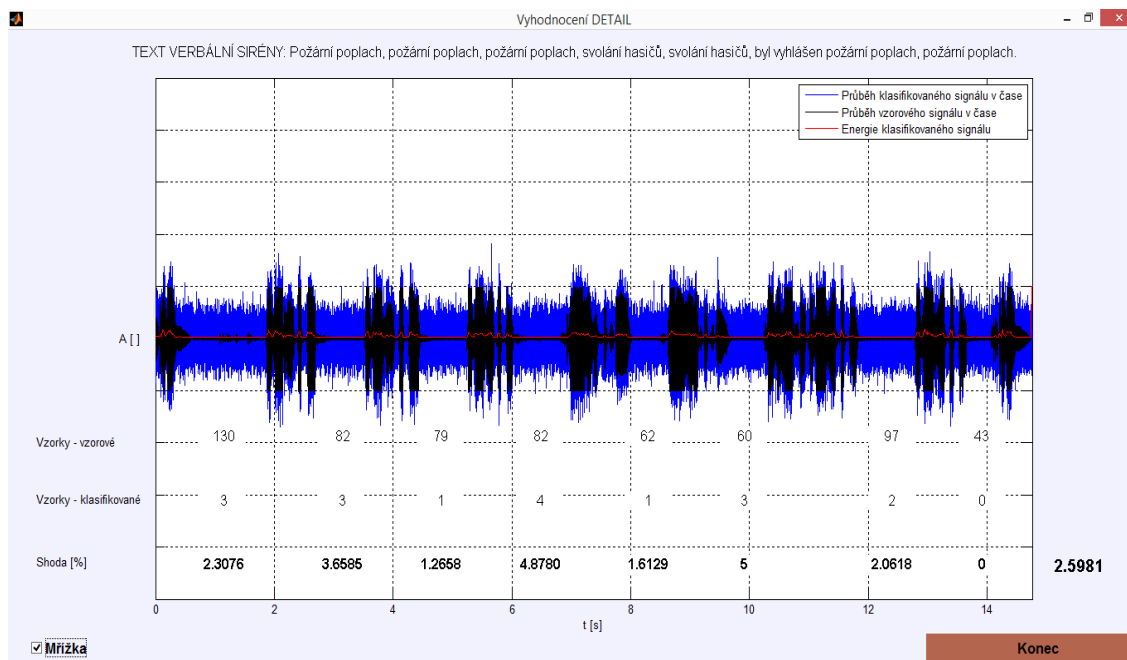
VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: + rušení (0.25)						
	Nebezpečí						
	Zkouška sirén	Požární poplach	zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	12.3072	4.8701	8.9550	24.3916	4.4404	2.1580	12.0482

Požární poplach	4.7749	6.0604	6.6271	29.0648	3.7991	3.3124	13.7301
Nebezpečí zátopové vlny	4.9025	2.8422	15.4401	26.0778	6.1164	2.8395	10.2450
Všeobecná výstraha	4.1339	2.7755	7.1525	56.9055	7.5428	2.0968	8.9525
Chemická havárie	4.1530	2.1571	5.6053	26.2022	11.9803	2.4439	12.9526
Radiační havárie	3.4174	2.1124	6.3614	25.8787	6.1299	6.7517	9.5002
Konec poplachu	2.7124	2.3832	3.8706	14.8550	4.0015	2.1866	17.6131

Matlab:

```
ruseni = random('normal', 0, 0.25, N, 1);
```

```
data = data_vzor + ruseni;
```



Obr. 46 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

Signály rušené přičtením náhodných dat o směrodatné odchylce 0,2 a méně, není problém klasifikovat s přesností $\pm 95\%$ i při zesílení či zeslabení signálů.

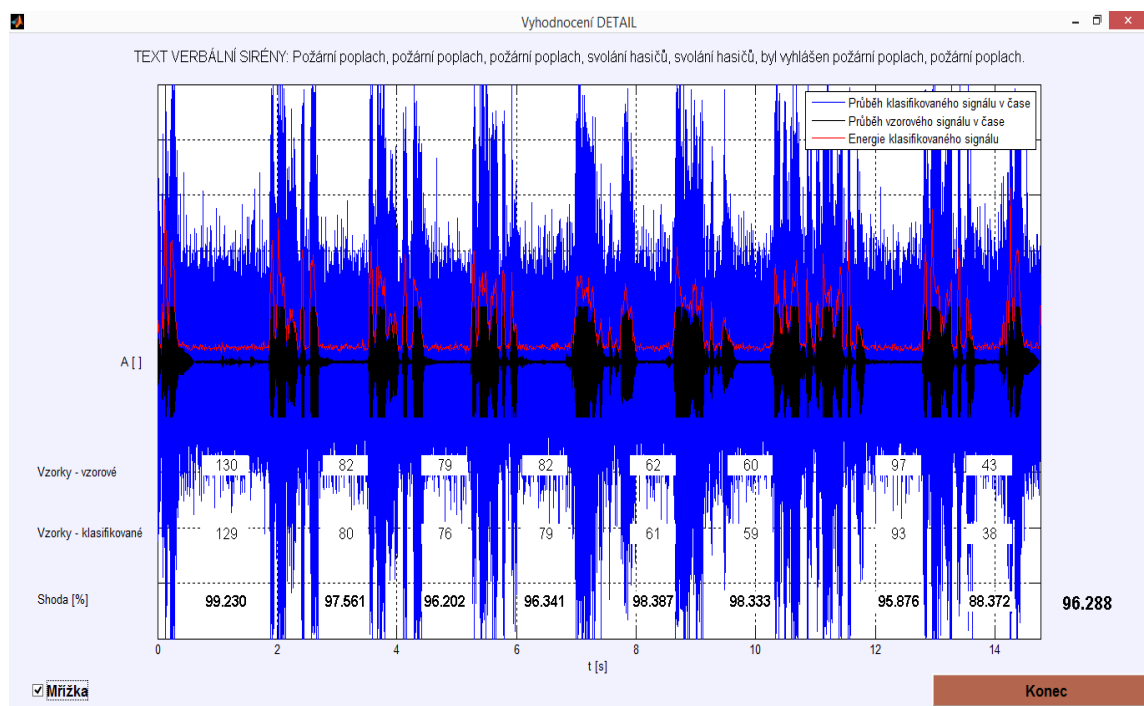
Tbl. 11 Výsledky klasifikace modifikovaného signálu (rušení, zesílení 4 krát) v [%].

VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: + rušení (0.2), zesílení *4						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	100	58.5023	57.3020	49.2441	48.1216	42.0613	62.0550
Požární poplach	61.1939	93.1765	42.8258	62.9771	40.9429	46.9747	73.6604
Nebezpečí zátopové vlny	55.1849	39.7347	99.3643	50.7196	44.2357	45.1921	53.9895
Všeobecná výstraha	39.7303	55.4049	47.9302	100	56.1247	52.1346	53.7966
Chemická havárie	39.1011	34.1100	33.7862	55.4858	88.9103	42.5626	71.2562
Radiační havárie	36.3879	40.4092	40.8619	53.9854	41.5840	90.7173	58.5423
Konec poplachu	29.8839	34.6839	23.3863	31.7474	32.0385	26.5146	90.5004

Matlab:

```
ruseni = random('normal', 0, 0.2, N, 1);
```

```
data = 4 * (data_vzor + ruseni);
```



Obr. 47 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

Z níže uvedených výsledků měření, lze snadno usoudit, že **mírné rušení** a zeslabení nemá vliv na důvěryhodnou klasifikaci.

Tbl. 12 Výsledky klasifikace modifikovaného signálu (rušení, zeslabení 0,2 krát) v [%].

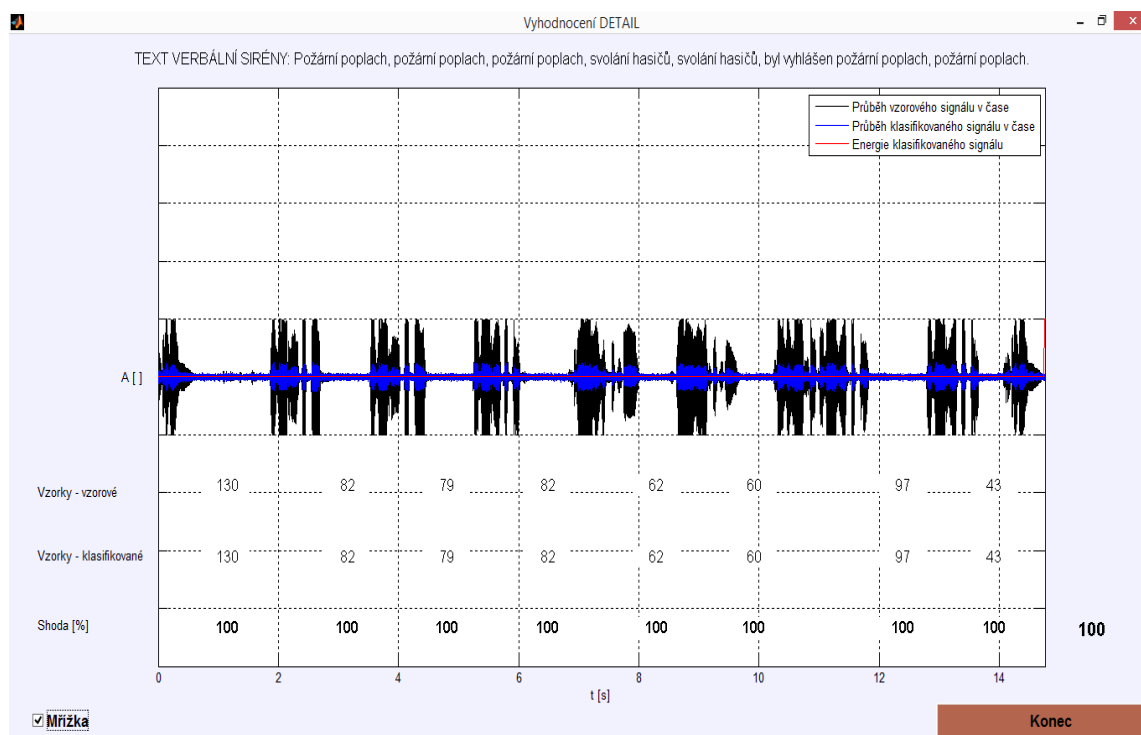
VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: + rušení (0.1), zeslabení *0.2						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	100	67.4138	68.8217	55.7622	64.3832	65.3175	79.1938
Požární poplach	70.0327	100	54.0260	68.4397	57.9737	65.0662	88.0236
Nebezpečí zátopové vlny	64.1814	47.2826	100	59.4564	62.6741	68.5915	73.2321
Všeobecná výstraha	52.9553	67.6119	57.9121	100	74.8653	73.1008	70.4817

Chemická havárie	48.5756	42.4826	43.0237	61.7214	100	65.2402	84.4373
Radiační havárie	44.7368	47.8271	52.1570	61.6807	56.2200	100	73.6829
Konec poplachu	34.5350	39.6652	29.3320	35.7704	44.6856	39.6887	100

Matlab:

```
ruseni = random('normal', 0, 0.1, N, 1);
```

```
data = 0.2 * (data_vzor + ruseni);
```



Obr. 48 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

8.3 Vliv posunu klasifikovaného signálu podél časové osy na detekci

Z výsledků měření, jež jsou uvedeny v tabulkách níže, je patrné, že **minimální posun** signálů nijak extrémně detekci neovlivňuje. Signály jsou klasifikovány s průměrnou přesností $\pm 97\%$.

Tbl. 13 Výsledky klasifikace modifikovaného signálu (posun, rušení, zesílení 1,5 krát) v [%].

VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: posun - 500 vzorků, + rušení (0.1), zesílení *1.5						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	98.3647	67.5027	66.4375	59.5111	67.3595	55.5675	80.6406
Požární poplach	70.0226	97.9804	52.3870	69.1025	62.0033	63.5072	80.8233
Nebezpečí zátopové vlny	64.6886	47.6190	96.4064	62.5771	61.2906	64.8807	72.1881
Všeobecná výstraha	50.9575	65.8416	57.6890	98.4329	78.7874	74.8014	68.2819
Chemická havárie	45.8898	43.6494	44.9342	62.4143	96.5474	63.6189	84.0851
Radiační havárie	45.4221	44.6257	53.3589	59.8199	57.3897	96.5304	71.5506
Konec poplachu	33.3468	40.8246	29.6074	36.6436	47.4768	39.7479	96.7103

Matlab:

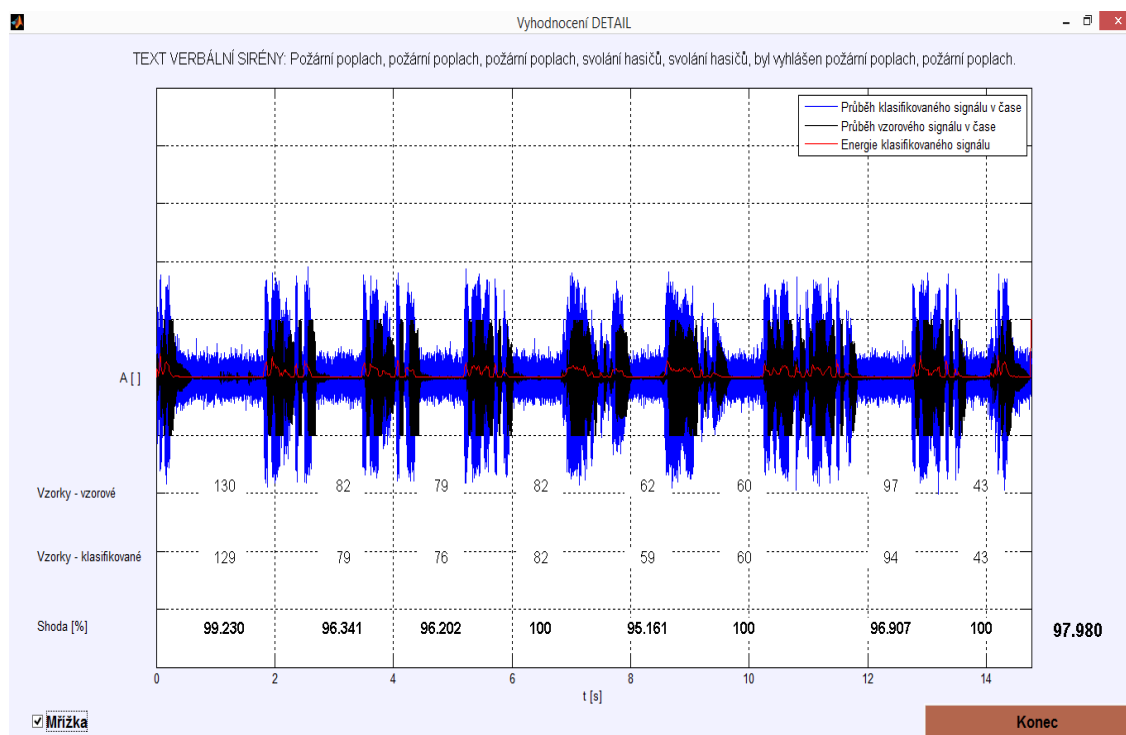
```
data_vzor_posunute = data_vzor(500:N);
```

```
data_vzor_posun = [data_vzor_posunute' zeros(1, N)];
```

```
data_posun = (data_vzor_posun(1 : N))';
```

```
ruseni = random('normal', 0, 0.1, N, 1);
```

```
data = 1.5 * (data_posun + ruseni);
```



Obr. 49 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

Razantnější posun už má viditelný vliv na procentuální shodu klasifikovaných signálů. Samozřejmě záleží také na jiných aspektech, které daný signál ovlivňují. V našem případě se jedná navíc ještě o rušení.

Tbl. 14 Výsledky klasifikace modifikovaného signálu (posun, rušení) v [%].

VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: posun - 1300 vzorků, + rušení (0.2)						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	85.2084	59.5880	55.5845	53.9576	40.1033	26.8985	59.6115
Požární poplach	61.3788	79.5736	37.2541	62.6218	31.7027	35.7930	56.5609
Nebezpečí zátopové vlny	56.8829	42.5509	69.7281	48.7052	27.0337	34.2441	45.7014

Všeobecná výstraha	40.3252	51.9687	46.1581	83.0681	45.1841	45.0834	45.7114
Chemická havárie	34.4613	35.7745	40.2037	57.3484	54.2880	31.0256	58.8124
Radiační havárie	41.3478	33.2758	43.6202	51.1129	25.8171	47.7688	44.6863
Konec poplachu	27.3625	36.5473	22.5656	34.3962	28.4684	21.2285	65.2311

Matlab:

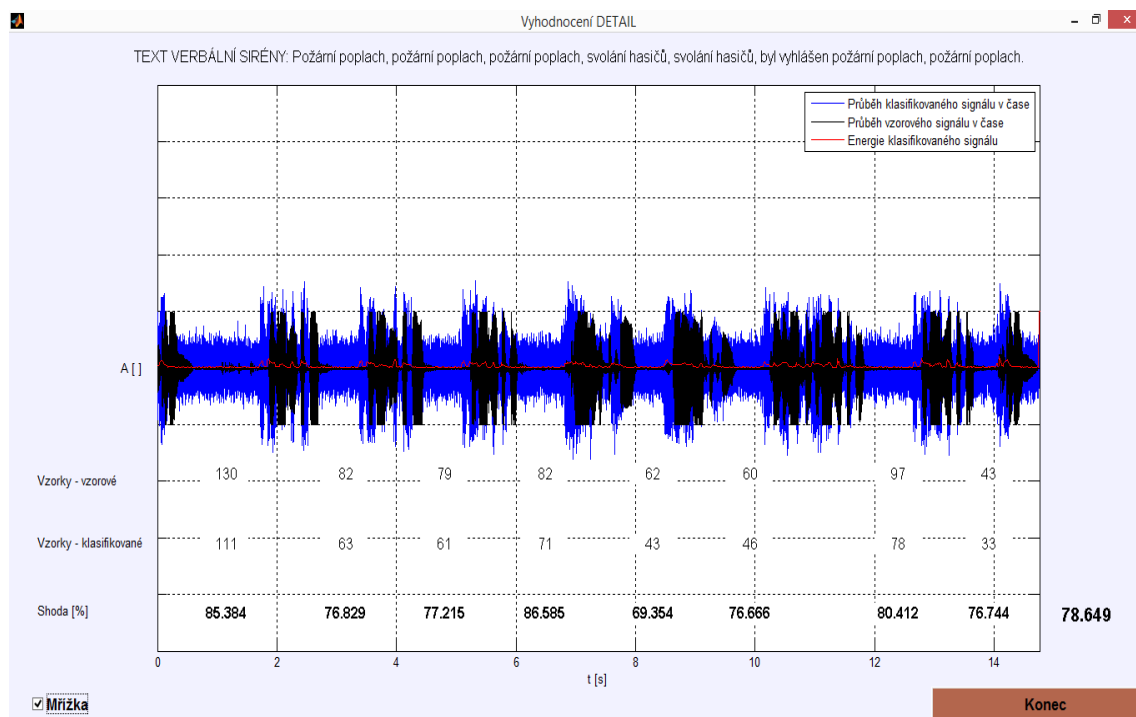
```
data_vzor_posunute = data_vzor(1300:N);
```

```
data_vzor_posun = [data_vzor_posunute' zeros(1, N)];
```

```
data_posun = (data_vzor_posun(1 : N))';
```

```
ruseni = random('normal', 0, 0.2, N, 1);
```

```
data = data_posun + ruseni;
```



Obr. 50 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

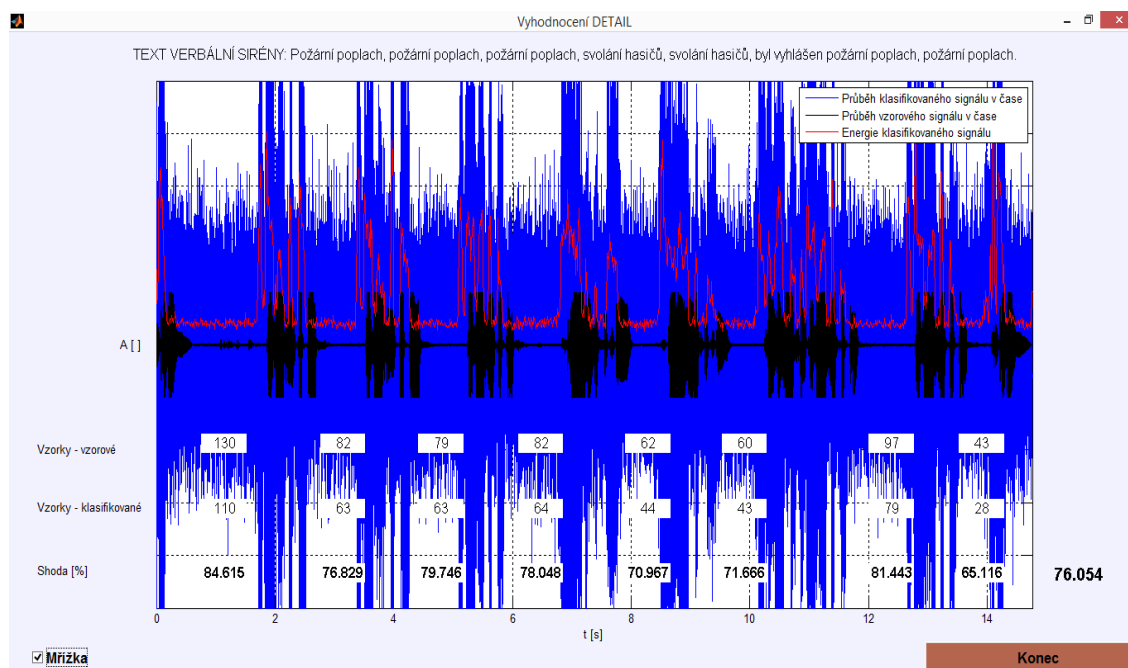
Níže je znázorněn totožný případ, jako v předchozím měření, jen doplněn o zesílení signálů 5 krát. Při porovnání výsledků s předchozím měřením vidíme chybu $\pm 4\%$, což nás utvrzuje, že zesílení signálů opravdu nemá extrémní vliv na finální výsledek měření.

Tbl. 15 Výsledky klasifikace modifikovaného signálu (posun, rušení, zesílení 5 krát) v [%].

VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: posun -1300 vzorků, + rušení (0.2), zesílení *5						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	85.4610	57.6400	54.5294	52.0549	43.3345	43.9899	75.2241
Požární poplach	60.5896	72.0394	37.5734	61.2821	35.3097	58.0111	71.3757
Nebezpečí zátopové vlny	58.2271	40.2778	68.9392	45.0737	30.1337	54.5154	60.8385
Všeobecná výstraha	40.6022	45.7389	44.8850	81.4830	52.2716	67.9477	59.4634
Chemická havárie	34.7379	31.7756	39.6437	54.3188	55.9459	49.2610	72.4053
Radiační havárie	42.6095	28.5742	44.4137	48.0354	30.4072	71.7020	57.8915
Konec poplachu	27.6677	33.1861	22.5220	33.4941	27.6367	34.2576	78.1108

Matlab:

```
data_vzor_posunute = data_vzor(1300:N);
data_vzor_posun = [data_vzor_posunute' zeros(1, N)];
data_posun = (data_vzor_posun(1 : N))';
ruseni = random('normal', 0, 0.2, N, 1);
data = 5 * (data_posun + ruseni);
```



Obr. 51 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

Od posunu -1500 vzorku byly identifikovány problémy s klasifikací některých signálů verbálních varovných sirén. Z měření níže vidíme, že nelze klasifikovat modifikovaný signál verbální sirény „radiacni_havarie.wav“.

Tbl. 16 Výsledky klasifikace modifikovaného signálu (posun, rušení, zeslabení 0,3 krát) v [%].

VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: posun - 1500 vzorků, + rušení (0.1), zeslabení *0.3						
	Nebezpečí						
	Zkouška sirén	Požární poplach	zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	88.5865	70.5328	65.0933	63.5666	68.1253	Nelze klasifikovat	83.5908
Požární poplach	70.8287	85.8097	55.6127	69.7224	64.5950	Nelze klasifikovat	80.5428
Nebezpečí zátopové vlny	62.2790	58.1540	79.6583	58.2540	58.5310	Nelze klasifikovat	72.0033

Všeobecná výstraha	54.9565	64.5771	58.9890	82.9063	83.0320	Nelze klasifikovat	71.0687
Chemická havárie	41.8312	44.2659	50.4437	62.3103	85.7103	Nelze klasifikovat	78.8467
Radiační havárie	49.0689	39.3978	61.1439	54.7540	54.0605	Nelze klasifikovat	69.8015
Konec poplachu	31.5433	39.8052	31.2335	39.5223	47.9986	Nelze klasifikovat	84.3588

Matlab:

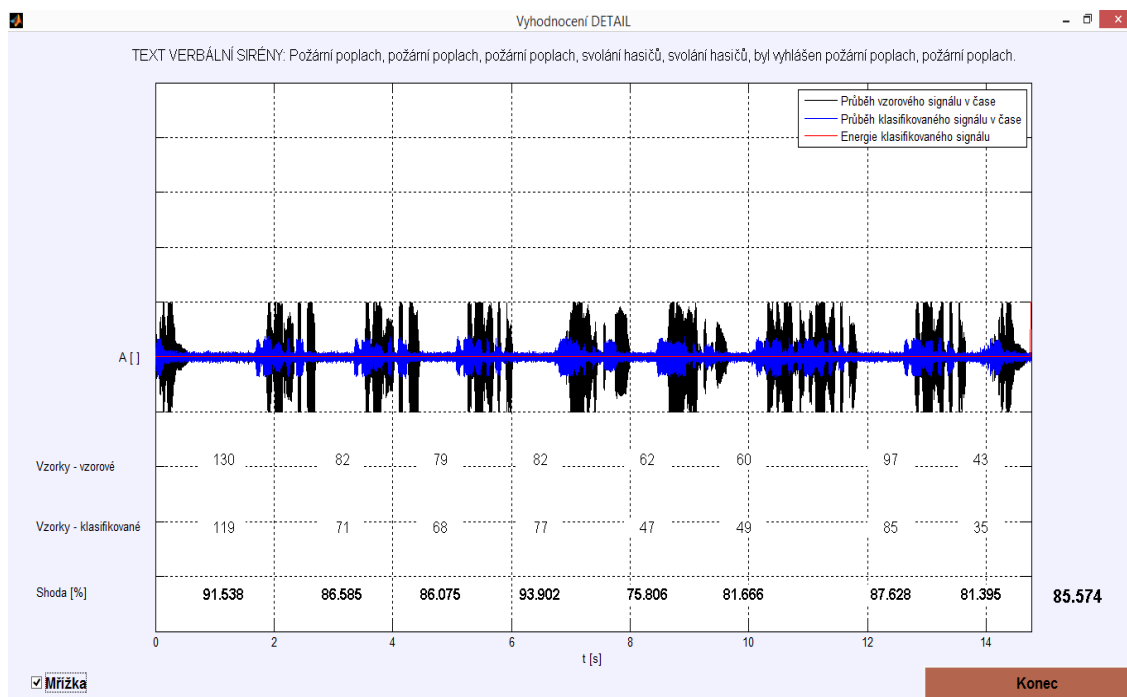
```
data_vzor_posunute = data_vzor(1500:N);
```

```
data_vzor_posun = [data_vzor_posunute' zeros(1, N)];
```

```
data_posun = (data_vzor_posun(1 : N))';
```

```
ruseni = random('normal', 0, 0.1, N, 1);
```

```
data = 0.3 * (data_posun + ruseni);
```



Obr. 52 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

Měření bylo uskutečněno také pro slabě rušený, zesílený signál s **posunem + 1000 vzorků**. Shoda signálů se pohybovala v průměru kolem $\pm 93\%$.

Tbl. 17 Výsledky klasifikace modifikovaného signálu (posun, rušení, zesílení 2 krát) v [%].

VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: posun + 1000 vzorků, + rušení (0.1), zesílení *2						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	94.0860	72.4298	71.7314	54.2499	58.1910	64.9998	85.0944
Požární poplach	69.5161	94.3475	58.3501	67.0081	53.6463	60.3105	87.9134
Nebezpečí zátopové vlny	64.6251	42.0294	90.7570	61.4099	68.2988	65.9958	75.0994
Všeobecná výstraha	52.0163	68.2414	54.0688	91.5171	70.1850	68.4564	74.8037
Chemická havárie	51.0039	48.4369	40.4558	64.5929	93.7731	67.6087	82.7778
Radiační havárie	41.8651	54.0645	51.1959	64.4762	58.6969	94.8000	80.7756
Konec poplachu	32.7171	36.4699	29.7722	34.1543	41.0573	39.4682	92.7122

Matlab:

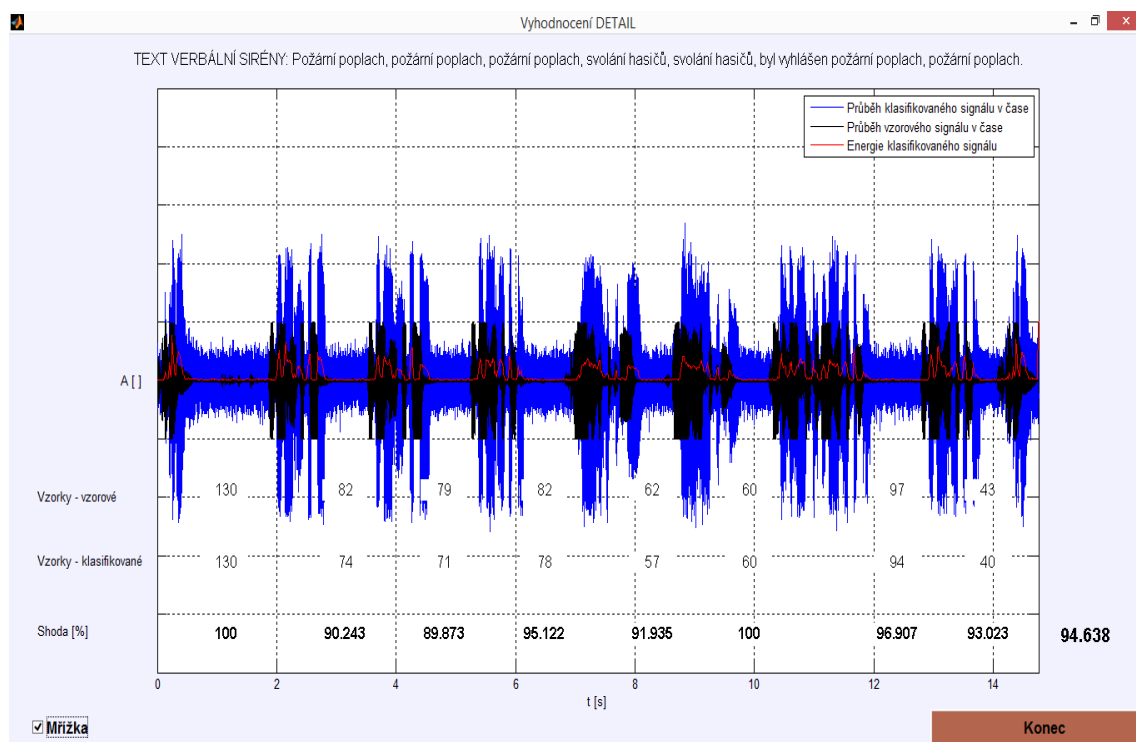
```
data_vzor_posun = [zeros(1, 1000) data_vzor'];
```

```
data_posun = data_vzor_posun(1 : N);
```

```
data_posun = data_posun';
```

```
ruseni = random('normal', 0, 0.1, N, 1);
```

```
data = 2 * (data_posun + ruseni);
```



Obr. 53 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

Při posunu + 1300 vzorků, zanedbatelném rušení a zesílení 7 krát mají signály průměrnou procentuální shodu $\pm 90\%$. Jak můžeme vidět z výsledků měření níže.

Tbl. 18 Výsledky klasifikace modifikovaného signálu (posun, rušení, zesílení 7 krát) v [%].

VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: posun + 1300 vzorků, + rušení (0.05), zesílení *7						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	90.0934	73.3804	71.7058	55.9734	57.9392	64.8245	83.0929
Požární poplach	70.5117	92.9168	61.7054	67.2900	53.8601	59.7008	87.6850
Nebezpečí zátopové vlny	62.2644	45.0910	86.1885	61.7139	71.7840	66.0832	76.5349

Všeobecná výstraha	55.1419	71.0873	54.6325	87.2569	72.2605	67.7942	78.0511
Chemická havárie	51.9646	50.2179	40.8577	83.8175	91.2698	69.5097	83.6764
Radiační havárie	41.9765	55.7280	53.3882	82.2637	60.0863	90.5175	83.0131
Konec poplachu	32.1305	35.8119	30.4655	78.6776	41.9237	39.4703	87.9760

Matlab:

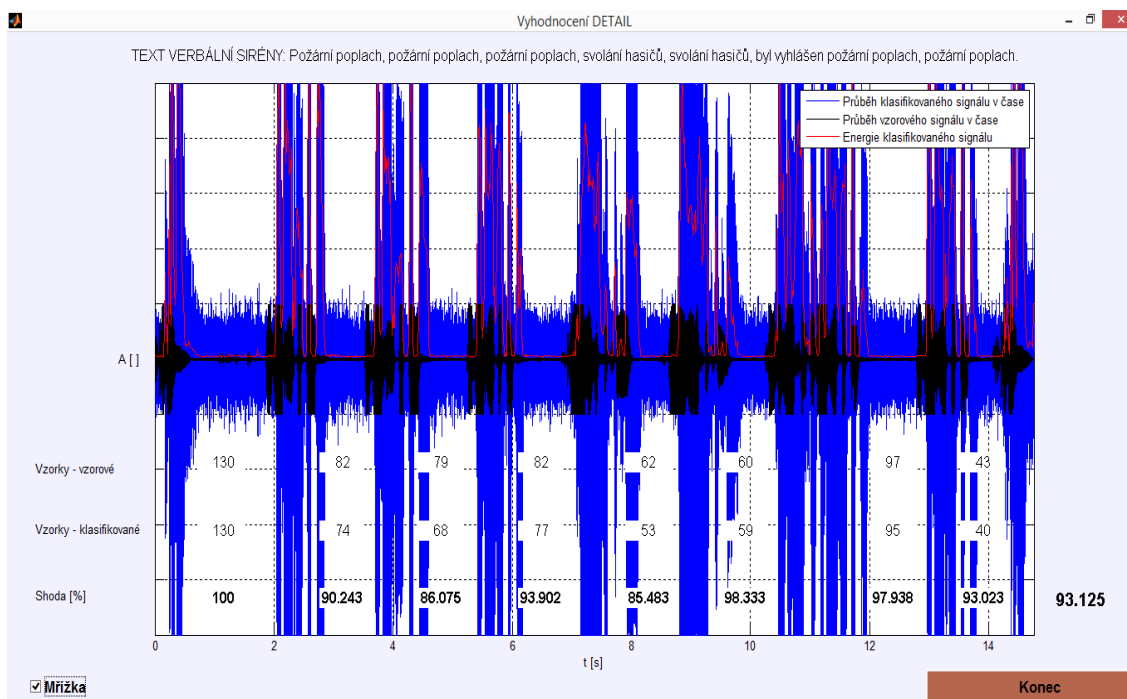
```
data_vzor_posun = [zeros(1, 1300) data_vzor'];
```

```
data_posun = data_vzor_posun(1 : N);
```

```
data_posun = data_posun';
```

```
ruseni = random('normal', 0, 0.05, N, 1);
```

```
data = 7 * (data_posun + ruseni);
```



Obr. 54 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

Při **posunu + 1800 vzorků**, rušení a zeslabení 0,1 krát, nebylo možno klasifikovat signál verbální sirény „konec_poplachu.wav“.

Tbl. 19 Výsledky klasifikace modifikovaného signálu (posun, rušení, zeslabení 0,1 krát) v [%].

VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: posun + 1800 vzorků, + rušení (0.2), zeslabení *0.1						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	72.1936	64.1673	56.8354	48.7014	41.2164	52.4157	Nelze klasifikovat
Požární poplach	64.2264	81.9839	51.1311	57.9035	38.0456	41.6388	Nelze klasifikovat
Nebezpečí zátopové vlny	53.5320	35.0503	65.6731	52.8081	62.1482	50.4524	Nelze klasifikovat
Všeobecná výstraha	52.6636	65.4562	34.7341	75.9075	53.7481	47.6869	Nelze klasifikovat
Chemická havárie	45.2987	45.6487	27.1368	59.5371	68.5415	49.3711	Nelze klasifikovat
Radiační havárie	34.6605	51.9499	40.5382	58.4516	48.5040	68.5075	Nelze klasifikovat
Konec poplachu	29.3695	31.5528	23.7968	27.9170	31.8469	25.3376	Nelze klasifikovat

Matlab:

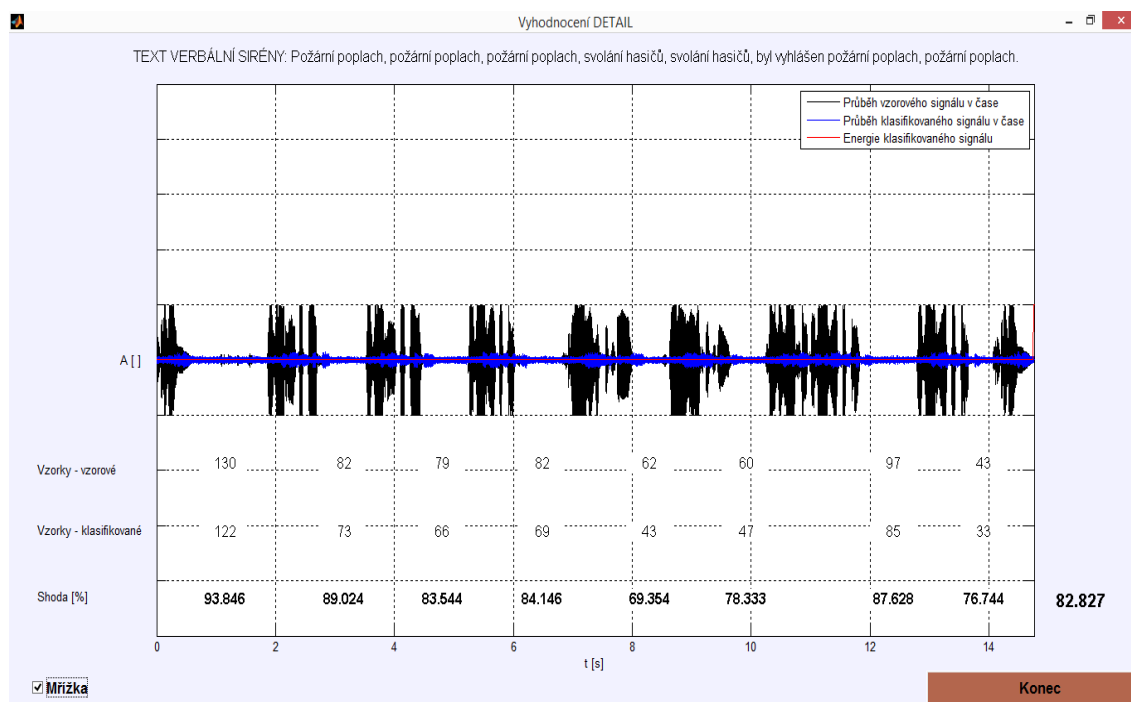
```
data_vzor_posun = [zeros(1, 1800) data_vzor'];
```

```
data_posun = data_vzor_posun(1 : N);
```

```
data_posun = data_posun';
```

```
ruseni = random('normal', 0, 0.2, N, 1);
```

```
data = 0.1 * (data_posun + ruseni);
```

Obr. 55 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

Ovšem při **posunu + 1800 vzorků** a slabém rušení, šly všechny signály klasifikovat bez problémů s průměrnou přesností $\pm 80\%$.

Tbl. 20 Výsledky klasifikace modifikovaného signálu (posun, rušení) v [%].

VZOROVÉ SIGNÁLY:	Modifikované signály dle algoritmu: posun + 1800 vzorků, + rušení (0.1)						
	Zkouška sirén	Požární poplach	Nebezpečí zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zkouška sirén	79.8446	70.0859	68.1902	55.6870	56.2998	60.8391	74.0544
Požární poplach	68.7568	87.9013	60.4032	63.6302	50.9636	54.6519	85.8740
Nebezpečí zátopové vlny	59.8754	41.6528	78.3530	60.2537	72.7036	64.1991	72.6937
Všeobecná výstraha	58.0166	71.8287	50.9402	83.0655	66.8197	63.3807	78.9532

Chemická havárie	49.9323	50.1748	39.1353	64.4821	82.0036	67.6135	84.9925
Radiační havárie	39.8941	54.9386	51.0109	63.6529	60.7073	85.2083	79.3581
Konec poplachu	31.6878	34.2451	30.2316	32.5535	40.4787	38.2424	87.3008

Matlab:

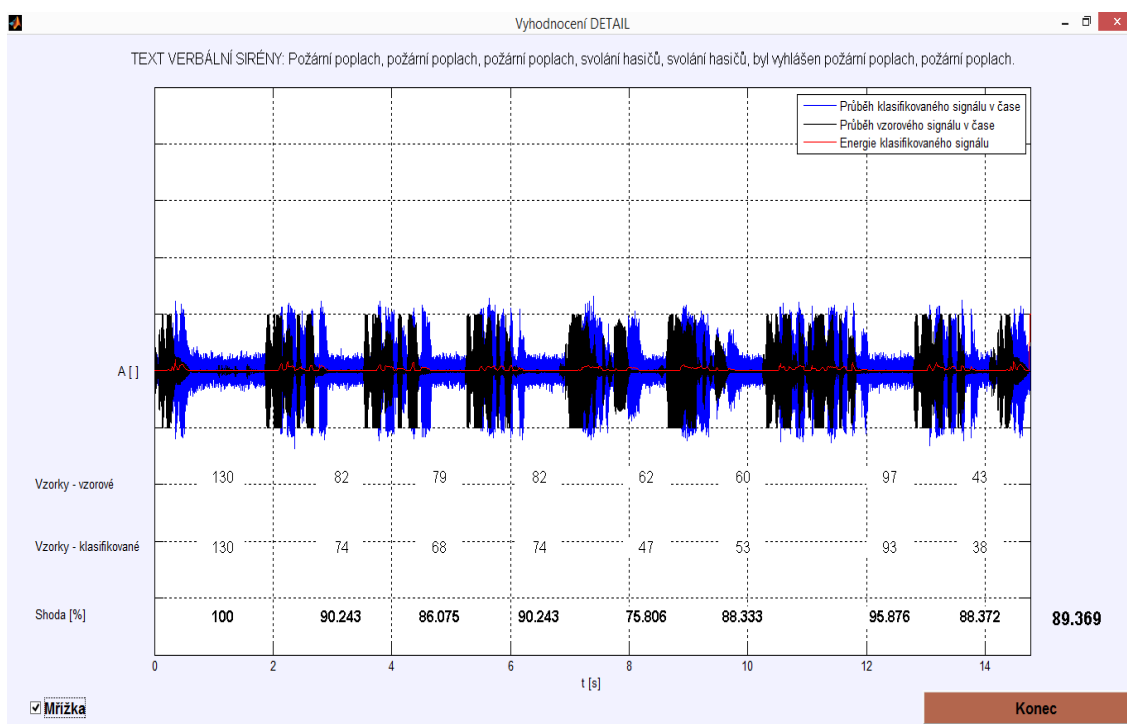
```
data_vzor_posun = [zeros(1, 1800) data_vzor'];
```

```
data_posun = data_vzor_posun(1 : N);
```

```
data_posun = data_posun';
```

```
ruseni = random('normal', 0, 0.1, N, 1);
```

```
data = data_posun + ruseni;
```



Obr. 56 Časový průběh vzorového signálu, energie a časový průběh klasifikovaného signálu požární poplach.

9. Zhodnocení dosažených výsledků

Z praktického měření, které bylo uskutečněno pomocí sestaveného softwarového programu lze snadno usoudit, že klasifikace jednotlivých signálů není nijak ovlivněna jejich samotným zesílením, či zeslabením. Samozřejmě v kombinaci s jinými abnormalitami (šum, posun) se projeví, nýbrž minimálně.

Naopak nežádoucí dopad na detekci by mohlo mít silné rušení, či razantní posun signálu podél časové osy. Z empirických zkušeností víme, že bylo problematické detekovat signály rušené přičtením náhodných dat se směrodatnou odchylkou větší než 0,25.

Co se týče posunu klasifikovaných signálů podél časové osy. První problémy s detekcí se začaly projevovat při samostatných posunech od -1900 vzorků k $-\infty$ a od +1900 vzorků k $+\infty$ a to přesně u modifikovaného signálu „konec_poplachu.wav“. V kombinaci s rušením se tyto problémy začaly vyskytovat při posunech menších.

Tabulka Tbl. 21 obsahuje **souhrn jednotlivých výsledků měření** v rámci klasifikace modifikovaného signálu. Vidíme zde procentuální shodu vzorového signálu se signálem modifikovaným dle předem stanovených algoritmů, jež jsou uvedeny v prvním levém sloupci tabulky. V případech, ve kterých je uvedena nula, nedošlo k rozpoznání klasifikovaného signálu. Tato situace nastala 2x.

Tbl. 21 Souhrn dosažených výsledků v rámci jednotlivých měření. Všechny výsledky jsou uvedeny v [%].

ALGORITMUS MODIFIKACE:	VERBÁLNÍ SÍŘENY:						
	Nebezpečí						
	Zkouška sirén	Požární poplach	zátopové vlny	Všeobecná výstraha	Chemická havárie	Radiační havárie	Konec poplachu
Zesílení: 20	100	100	100	100	100	100	100
Zeslabení: 0.01	100	100	100	100	100	100	100
Rušení: silné	12.3072	6.0604	15.4401	56.9055	11.9803	6.7517	17.6131
Rušení: střední, zesílení: 4	100	93.1765	99.3643	100	88.9103	90.7173	90.5004

Rušení: slabé, zeslabení: 0.2	100	100	100	100	100	100	100
Posun: - 500 vzorků, rušení: slabé, zesílení: 1.5	98.3647	97.9804	96.4064	98.4329	96.5474	96.5304	96.7103
Posun: - 1300 vzorků, rušení: střední	85.2084	79.5736	69.7281	83.0681	54.2880	47.7688	65.2311
Posun: -1300 vzorků, rušení: střední, zesílení: 5	85.4610	72.0394	68.9392	81.4830	55.9459	71.7020	78.1108
Posun: - 1500 vzorků, rušení: slabé, zeslabení: 0.3	88.5865	85.8097	79.6583	82.9063	85.7103	0	84.3588
Posun: + 1000 vzorků, rušení: slabé, zesílení 2	94.0860	94.3475	90.7570	91.5171	93.7731	94.8000	92.7122
Posun: + 1300 vzorků, rušení: zanedbatelné, zesílení: 7	90.0934	92.9168	86.1885	87.2569	91.2698	90.5175	87.9760
Posun: + 1800 vzorků, rušení: střední, zeslabení: 0.1	72.1936	81.9839	65.6731	75.9075	68.5415	68.5075	0
Posun: + 1800 vzorků, rušení: slabé	79.8446	87.9013	78.3530	83.0655	82.0036	85.2083	87.3008
PRŮMĚR:	85	84	80	88	79	73	77

Poznámky:

Zesílení: zesílení signálu

Zeslabení: zeslabení signálu

Posun: posun signálu podél časové osy

Rušení: zanedbatelné (přičtení dat se směrodatnou odchylkou 0,05),

slabé (přičtení dat se směrodatnou odchylkou 0,1),

střední (přičtení dat se směrodatnou odchylkou 0,2),

silné (přičtení dat se směrodatnou odchylkou 0,25)

10. Závěr

V rámci diplomové práce jsem se zaměřila na teoretický rozbor problematiky předzpracování, zpracování a detekce mluveného signálu s praktickou implementací pomocí programu Matlab. Dále jsem navrhla a realizovala celistvý software pro detekci verbálního signálu s následnou vizualizací a srovnáním naměřených výsledků.

Softwarovou aplikaci jsem vyvinula v prostředí **Matlab**, je schopná **detekovat jednotlivé elektronické verbální sirény**, sloužící k varování obyvatelstva před hrozbou nebo při vzniku mimořádné situace. Jedná se o sirény - zkouška sirén, požární poplach, nebezpečí zátopové vlny, všeobecná výstraha, chemická havárie, radiační havárie, konec poplachu.

Samotný **princip realizace** tohoto detektoru **spočívá ve výpočtu energie** jednotlivých signálů a součtu vzorků v mezerách mezi jednotlivými souslovími, jejíž amplituda energie dosahuje nanejvýš 10% z maximální hodnoty. Na tomto základě dochází k porovnání klasifikovaného signálu se signály vzorovými, výpočtu procentuální shody a vypsání výsledku do informačního okna – „klasifikace signálu“.

V rámci **ověření vytvořeného softwaru** jsem uskutečnila **měření pomocí 13-ti algoritmů**, díky kterým jsem modifikovala jednotlivé vzorové signály. Jedná se o zesílení, zeslabení, posun, rušení a kombinaci všech. Takto upravené signály jsem poté detekovala vyvinutou softwarovou aplikací.

Z praktického měření lze vyhodnotit, že klasifikace jednotlivých signálů nebyla nijak ovlivněna jejich samotným **zesílením**, či **zeslabením**. Samozřejmě v kombinaci s jinými abnormalitami (šum, posun) se projevila, nýbrž minimálně.

Naopak nežádoucí dopad na detekci mělo **silné rušení**, či razantní **posun signálu podél časové osy**. Problematické bylo detekovat signály rušené přičtením náhodných dat se směrodatnou odchylkou vyšší než 0,25.

Co se týče posunu klasifikovaných signálů podél časové osy, první problémy s detekcí se začaly projevovat při samostatných posunech od -1900 vzorků k $-\infty$ a od +1900 vzorků k $+\infty$ a to přesně u modifikovaného signálu „konec_poplachu.wav“. V kombinaci s rušením se tyto problémy začaly vyskytovat již při posunech menších.

Průměrná procentuální shoda klasifikovaného signálu se signálem vzorovým, v rámci uskutečněných měření, dosahovala pro zkoušku sirén **85 %**, požární poplach **84 %**, nebezpečí zátopové vlny **80 %** všeobecnou výstrahu **88 %**, chemickou havárii **79 %**, radiační havárii **73 %** a konec poplachu **77 %**.

Literatura

- [1] **PSUTKA, Josef, a další.** *Mluvíme s počítačem česky*. Praha : Academia Praha, 2006. ISBN-80-200-1309-1.
- [2] **PSUTKA, Josef.** *Komunikace s počítačem mluvenou řečí*. Praha : Academia Praha, 1995. ISBN 80-200-0203-0.
- [3] **LAVICKÝ, Martin.** *Demonstrativní systém na rozpoznávání jednotlivých slov*. Diplomová práce. Brno: Fakulta elektrotechniky a informačních technologií VUT, 2003.
- [4] **ČERNOCKÝ, Jan.** *Zpracování řečových signálů - studijní opora*. [online]. [cit. 1. 11. 2014]. Dostupné z: <http://www.fit.vutbr.cz/study/courses/ZRE/public/opora/zre_opora.pdf>
- [5] *Leccos - Souhláska*. [online]. [cit. 1. 11. 2014]. Dostupné z: <<http://leccos.com/index.php/clanky/souhlaska>>
- [6] **MCLOUGHLIN, Ian.** *Applied Speech and Audio Processing*. Leiden: Cambridge University Press, 2009. 216 s. ISBN 978-0-521-51954-0.
- [7] **VONDRA, Martin.** *Kepstrální analýza řečového signálu*. Brno: Vysoké učení technické v Brně. [online]. [cit. 10. 11. 2014]. Dostupné z: <<http://www.elektrorevue.cz/clanky/01048/index.html>>
- [8] **TARANTINOV, J., POLLÁK, P.** *Řečové detektory využívající ergodické Markovovské modely*. Praha: České vysoké učení technické v Praze. [online]. [cit. 10. 11. 2014]. Dostupné z: <http://dsp.vscht.cz/konference_matlab/MATLAB08/prispevky/109_tatarinov.pdf>
- [9] **RAJNOHA, J., POLLÁK, P.** *Detektory řečové aktivity na bázi perceptivní kepstrální analýzy*. Praha: České vysoké učení technické v Praze. [online]. [cit. 11. 11. 2014]. Dostupné z: <http://dsp.vscht.cz/konference_matlab/MATLAB08/prispevky/088_rajnoha.pdf>
- [10] **PORUBA, Jiří, MATĚJÍČEK, Lukáš.** *Odfiltrování rušivých signálů ze zašumělé řeči*. Článek Elektrorevue. Brno: Vysoké učení technické v Brně, 2002. [online]. [cit. 20. 11. 2014]. Dostupné z: <<http://www.elektrorevue.cz/clanky/02047/index.html>>
- [11] **VONDRÁČEK, M.** *Odhad SNR řečového signálu snímaného v hlučném prostředí*. Diplomová práce. Praha: České vysoké učení technické v Praze, 2004.
- [12] *Požáry*. [online]. [cit. 20. 1. 2015]. Dostupné z: < <http://www.pozary.cz/clanek/54244-system-jednotneho-varovani-a-vyrozeni-v-ceske-republice/> >
- [13] *Stavba a funkce hlasového ústrojí*. [online]. [cit. 20. 1. 2015]. Dostupné z: <<http://pfyziolfup.upol.cz/castwiki/?p=2661>>

[14] *Diskrétní signál*. [online]. [cit. 20. 3. 2015]. Dostupné z: http://midas.uamt.feec.vutbr.cz/ZVS/Exercise01/content_cz.php

[15] **BARDICI, Nick, a další.** *Speech Recognition using Hidden Markov Model*. School of Engineering: Department of Telecommunications and Signal Processing, 2006.

Seznam příloh

I. Příloha na CD

Matlab

- frame.m
- frame_2.m
- dft.m
- energie.m
- mean.m
- Z_E_M.m
- klasifikace_signalu.m
- mezery_1.m
- mezery_2.m
- mezery_3.m
- mezery_4.m
- mezery_5.m
- mezery_6.m
- mezery_7.m

Verbalni_sireny

- chemicka_havarie.wav
- konec_poplachu.wav
- nebezpeci_zatopove_vlny.wav
- pozarni_poplach.wav
- radiacni_havarie.wav
- vseobecna_vystraha.wav
- zkouska_siren.wav

HOL0132_Rozpoznavaci_system_pro_detekci_mluveneho_signalu.pdf